

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

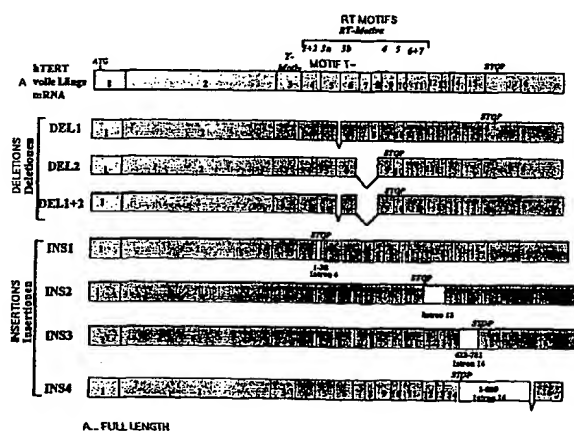


INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998
		(43) Internationales Veröffentlichungsdatum:	8. Juli 1999 (08.07.99)
(21) Internationales Aktenzeichen: PCT/EP98/08216		(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)		<p>Veröffentlicht</p> <p><i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>	
(30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE			
(71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).			
(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).			
(74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).			

(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF

(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshjan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland		Republik Mazedonien	TR	Türkei
BG	Bulgarien	HU	Ungarn	ML	Mali	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MN	Mongolei	UA	Ukraine
BR	Brasilien	IL	Israel	MR	Mauretanien	UG	Uganda
BY	Belarus	IS	Island	MW	Malawi	US	Vereinigte Staaten von
CA	Kanada	IT	Italien	MX	Mexiko		Amerika
CF	Zentralafrikanische Republik	JP	Japan	NE	Niger	UZ	Usbekistan
CG	Kongo	KE	Kenia	NL	Niederlande	VN	Vietnam
CH	Schweiz	KG	Kirgisistan	NO	Norwegen	YU	Jugoslawien
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik	NZ	Neuseeland	ZW	Zimbabwe
CM	Kamerun		Korea	PL	Polen		
CN	China	KR	Republik Korea	PT	Portugal		
CU	Kuba	KZ	Kasachstan	RO	Rumänien		
CZ	Tschechische Republik	LC	St. Lucia	RU	Russische Föderation		
DE	Deutschland	LI	Liechtenstein	SD	Sudan		
DK	Dänemark	LK	Sri Lanka	SE	Schweden		
EE	Estland	LR	Liberia	SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der
30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischem Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus
15 von Tumorzellen zur Stabilisation ihrer Telomere.

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzwelligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

5
10
15
20
25
30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

5 RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

10 Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

15 Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

30 Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.
15

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-
30

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (IGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

30

- A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.
- Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.
- Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.
- Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.
- Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergen-konstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

- 20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
20 14: nicht verdaut.

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

10

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

30

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

30

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

15

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind

30

gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

- 5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
- 10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Exon	Intron	Intron on	bp	
						No.		No.
5' flanking Region								
caggcgcttccccgtag	GTTCAGGCAGCGCTGCGT	1	281	CGCCCCCTCCTTCGCCAG	gtggcctccccgggtcg	1	104	
catgtccttctcgtttaag	GTGTCCTGCTGAAGGAGC	2	1354	TGGCTGGCAGGAGCCAG	gtgaggaggtggtggcgt	2	8616	
gaggggtctctattgtag	GGGTGGCTGTGTTCCGGC	3	196	TGCAAGCATTTGGAATCAG	gtactgtatccccacgcca	3	2089	
cccatgctgtccccgtag	ACAGCACTTGAAGAGGGTG	4	181	GTTCGCCAGAGAAAAAGAGG	gtggctgtgctttggttta	4	687	
ctcgctcccaactcacatag	GCCGAGCGTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgccccgggaccccc	5	494	
ccctctcctctgcccggtag	GTGATGTGACGGGGCGGT	6	156	CAAGGCCTTCAAGAGCCAC	gtagggttcacgtgtgata	6	>4660	18
ctccgctctgcttctgtag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGTCATCGAGCAG	gtctgggcaactgccccgca	7	980	
ctgtgtcttcccgcccttag	AGCTCCTCCCTGAAATGAGG	8	86	CCGTGCCCATCAGGGGCAA	gtgagtcagggtggcccaggt	8	2484	
gtattttcccttatttttag	GTCTACGTCCAGTGCCAG	9	114	CGGGGATTCCGGCGGACGG	gtgaggcctcctcttcccc	9	1984	
cattgccccctctgccttag	GCTGCTCCTGCGTTTGGTG	10	72	ACGCGAAAAACCTTCCTCAG	gtgagggccccgtccgtgtg	10	1871	
attccccctgtgtctdag	GACCTGGTCCGAGGTGTC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgcaactggcccggga	11	3804	
tcctttttggcgactcttag	CTATGCCCGGACCTCCATC	12	127	CCTGTTTCTGGATTTCAG	gtgagcaggctgatggtea	12	880	
ctgtccgccaactctctdag	GTGAACAGCCTCCAGACGG	13	62	TCCTGTGTCAGGCGTACAG	gtgagccgccaccaagggg	13	3184	
agcctctgttttccccdag	GTTTACCGCATGTGTGCTG	14	125	CTGAAAGCCAAAGACGCAG	gtatgtgcagggtgcctggc	14	781	
tctgattttggccccgtag	GGATGTGCTGGGGGCCAA	15	138	CTGGGGTCACTCAGGACAG	gcaagtgtggtggaggcc	15	536	
	CCGACAGCGCAGCTGATCG	16	664	TTTTTCAGTTTTTGAAAAAA	3' flanking Region			

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

- Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.
- Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.
- Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.
- Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

15 ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20 kloniert.

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses
Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen
genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche
20 identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus
Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic
Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5
zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG
Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2
unterstrichen.

Contig1:

5 ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA 70
 ATGAGACCCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGA 140
 ACAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAACT ATACAAACAC ATGAAATTA ACAATATAC 210
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAGGAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA 280
 CGGAAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA 350
 AGCAGCTACA TCAAAAAAGT AGAAAAAGCA GGCGCAGTGG CTGATGCCTG TAATCCCAGC ACTTTGGGAG 420
 10 GCCAAGGCGG GCAGATCGCC TGAGGTCCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTGC 490
 CTACTAAAAA TACAAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG 560
 GCAGGATAAC CGCTTGAAAC CAGGAGGTGG AGGTTGCCGT GAGCCGGGAT TGGCCCATTT GACTCCAGCC 630
 TGGGTAAACAA GAGTGAACCC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT 700
 TAGTGCACCTT AAAGAAGTAG AAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAGAAA AGAAATATAA 770
 AAGATCAGAG CAGAAATAAA TGAACCTGAA AGATAACAAT ACAAAGATC AACAAAAATTA AAAGTTGGTT 840
 15 TTTTGAAGAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA 910
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACGTATAC CACAGAAAT CAAAGGATCA CTAGAGGCTA 980
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTAGA TGCATACAC 1050
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAA GACCAATAAC AATATGGGA TTAAGGCCAT 1120
 20 AATAAAGT CTCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT 1190
 AAAGAAGAT GAATTCATC CTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT 1260
 TCTACATGGC CAGTATTACC CTGATTCCAA AACAGACAA AAACACATCA AAAACAAACA AAAAAAAA 1330
 CAGAAACAAA GAAACTACA GGCAATATC CTGATGAAT ACTGATACAA AAATCCTCAA CAAACACTAA 1400
 GCAACCAAAA TTAACAACA CCTTCGAAG ATCATTCTT GTGATCAAGT GGGATTTATT CCAGGGATGG 1470
 25 AAGGATGGTT GAGTATGAG AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAATAAT 1540
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTATGATA AAAACCTCA 1610
 AAAAAACGAG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGGCATCCCA GCACCTGGG 1680
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT 1750
 CTACAAAAAA CTTTTTAA AATTAGCCA GGCAATGAGT CATATGCCTG TAGTCCCAGC TAGTCTGGAG 1820
 30 GCTGAGGTGGT GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT 1890
 CCAGCCTAGA CAACAGACA AGACCCCACT GAATAAGAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG 1960
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG 2030
 AAGCAAGATA TTCAACATA ATAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGA 2100
 AGCCTTCTCT CTAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA 2170
 35 AGTCATTAGT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA 2240
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAGA GACTTAAGAC ACCACTAAAA AACTATTAGA 2310
 GCTGAAATTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCAAACAGC 2380
 AAACAATCTG AAAAAGAAAC CAAAAAGGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAGAAC 2450
 GTGAAAGATG TCTACAATGA AAATAATAAA ATGTTGATAA AAGAAATGA AGAGGGCACA AAAAAAGAAA 2520
 40 AGATATTCCA TTGTCATAGA TTGGAAGAAT AAATACTGTT AAATGTCCA TACTACCAA AGCAATTTAC 2590
 AAATTCAATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT 2660
 TTGTACAGAA CCACAAAAGA CCCAGAATAG CCAAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT 2730
 CACATTAAGT GACTTCAAT TATACTACAA AGCTATAGTA ACCCAACTA CATGGTACTG GCATAAAGAC 2800
 AGATGAGACA TGGACCAGAG GAACAGAAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA 2870
 45 TTTTGGACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTT AATAAATGGT GCTGGAGGAA 2940
 CTGGATATCC ATATGCAAAA TAACAATACT AGAAGCTCTG CTCTACCAT ATACAAAAGC AAATCAAAAT 3010
 GGTATGAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAGAA AACACCGGAG AAATCTCCA 3080
 GGCAATTGGA GTGGGCAAGG ACTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAGC AAAAAAGAC 3150
 AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCAGCAA AGGAAACAT CAACAAGAG AGAGACACA 3220
 50 CCACAGAAATG GGAGAAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATACCAAGTA TATATAAGGA 3290
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAAGTGAT TTTCAAAAT AAGCAAAAGA TCTGGGTAGA 3360
 CATTTCTCAA AATAAGTCAT ACAATAGGCA AACAGGCATC TGAAATGTG CTCACACCA CTGATCATCA 3430
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAGAC 3500
 AGGCAATAC AAATGCCAGT GAGGATGTGG ATAAAGGAA ACCCTGGAC ACTGTTGGTG GGAATGGAAA 3570
 55 TTGCTACCACT TATGGAGAAC AGTTTGAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC 3640
 CCATTGTCTAG GTATATACTC CAAAAAGGG AATCAGTGTA TCAACAGCT ATCTCCACT CCACATTTAC 3710
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG 3780
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA 3850
 CAGCATGGGG GGCACCTGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCAATGT 3920
 60 CTCCCTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG 3990
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTAA AATAACTAAA AGAGTATAAT 4060
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTAC CCTGATGTGA 4130
 TTATTACATA TTGTATGCC TATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCCTA CTATATTAAA 4200
 AATTAAAAAT TTAATGGCCA GGCACGGTGG CTGATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG 4270
 65 TGGATCACC TGAGGTCAGG AGTTTGAAC CAGCTGGCC ACCATGATGA AACCTGTCT CTACTAAGA 4340
 GTCAAAAAAT AGCCAGGCGT GTGGGCACAT ACCTGTAGTC CCACTACTC AGGAGGCTGA GACAGGAGAA 4410
 TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA 4480
 GAGCAAGACT CCATCTCAA ACAAAGAAAG ATTAATAATT TAATTTTAT TAATTTTAT GTACCGTATA 4550
 AATATATACT CTAATATATT AGAAGTTAAA AATTAACAA ATTATAAAG GTAATTAACC ACTTAATCTA 4620
 70 AATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAACT ATGGCCACCA TGGCAGAAAT 4690
 GTGAGGAGGG AACAGTGGA GTTACTGTTG TTAGAGCTCT ATACTCTCTG TAAGTGACTT AATTTTAACC 4760
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAAACA CTGCTAATAA TGGTGAAGG 4830
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA 4900
 CACCGTCTCT TCATTACGGG TGCTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTG 4970
 75 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGCT GGTGATTTCC TCCAGAAGAA TTAGAGTACC 5040
 TGGCAGGAAG CAGGTGGCTC GTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG 5110

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
TCGAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180								
CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAGAGATTG	CTCTGGATAC	CATCTGGAAA	5250								
AGGCGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320								
CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTACCTGA	5390								
AGCAGTGACC	AAGGTGTATT	CTGAGGGAAG	CTTGAGTTAG	GTGCCTTCTT	TAAACAGAA	AGTCATGGAA	5460								
GCACCCCTCT	CAAGGGAAAA	CCAGACGCCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTTG	5530								
CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCCGTGA	GCTTCTCCGA	GCCCGTGCTG	AGGACCTCT	5600								
TGCAAGAGGC	TCCACAGACC	CCCGCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACTGGGATG	5670								
TGGCTGGGGG	CGGACAGCGA	CGCGGGGATT	CAAGAGCTTA	ATTCCATGAG	TAAATCAAC	CTTTCCACAT	5740								
CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTGAGGA	CTGCAGAAAT	5810								
CCAAAGGCGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAGG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880								
ATTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTTATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950								
AGAGGCACTG	CGCGCAGGGC	TATGAGCACG	GCAGGGCCAC	CGGGGAGAGA	GTCCCGGGCC	TGGGAGGCTG	6020								
ACAGCAGGAC	CACTGACCGT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	GGAAGCGGCG	CACGCTGCGT	6090								
GTGACTCAGG	ACCCCATACC	GGCTTCCTGG	GGCCACCCAC	ACTAACCCAG	GAACTCACGG	AGCTCTGAAC	6160								
CCGTGGAAAC	AAACATGACC	CTTGCTGCCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	GTGGTGTGCA	6230								
GGAAATGGCC	ATGTAATAAT	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCATTATT	CATCTTACC	6300								
CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TGCGGTGTGA	CAAGCCATGA	CAAAACTCAG	TACAAACCTG	6370								
ACTCTTTTAC	TAGGCCACAC	GAGCAGGSC	CACACCCCTG	ATATATTAA	AGTCCAGGAG	AGATGAGGCT	6440								
GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGCTGTGTCC	TCTAGACTAG	TAGACCTTGG	6510								
CAGGCACCTC	CCGAGTTTCT	AGGGCTCTTC	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580								
CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCTCCAG	GCCTCAGCTT	CTCCAGCAGC	6650								
TTCTATAAAC	CTGGGTGGGG	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CACGTGTGCT	TGCTCTCAGC	6720								
ACGTAGCTCG	CACGGTTCCT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	CGCTTGAAGG	6790								
GAGGAGATTG	TGCGCTTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GATGCAGGTT	6860								
CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TCCAGGCGC	TCCCGCTCTC	CTGTCTCTG	CCGGGGCCTG	6930								
CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	CTAGGGTCTC	7000								
GGGTTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAATGCA	CATTGTTGGTG	7070								
TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGSCA	CAGGCCTGGG	GATGGAGCCC	CCGCCAGGGA	7140								
CCCGCCCTTC	TCTGCCCAGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TCCACAAGC	7210								
ACTAAGACTG	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CCTGGGAATT	7280								
CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGC	CTGTTTATT	TAAATAGCTA	7350								
CAAAAGCAGG	AAATCCCTGC	TAAATGTCC	TTTAAACAA	TGGTTAAACA	AACGGGTCCA	TCCGACCGGT	7420								
GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490								
GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TAGCACAACAT	GCTCAAAAG	AAAGAATTTC	ACCCCATGGC	7560								
AGGGGAGTGG	TTAGGGGGGT	TAAAGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTACTA	7630								
AAGCCAGTTT	CCTGGTTCG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700								
GGAAACCCGA	GAGCTGTGCA	TCTTTGCCAT	CGCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCCC	7770								
ACGTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCC	GCCCCAGGCG	CTTTGCAGGT	GTGATCTCCG	7840								
TGAGGACCTT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTCTGGGGA	7910								
AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCACTGTA	GCGTGAAGAG	7980								
GGAGGGAGGG	CCTCGAGCCC	AGGCCCTGCA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCTTCC	8050								
ACGGACCTGG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAAGGGCC	ATCGTGAGCC	TCCGGCCTCC	8120								
GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTGCAGAA	GCAACAGGAA	8190								
ACCCATGAC	TGTGAATCTA	GGATTATTTC	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260								
GTGCTTCGGG	GCAAGGGCAG	GGCAGGACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTT	8330								
CTGAGACAGA	GTTATGCTCT	TGTTGCCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400								
CGTCTCTTGA	GTACAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTACAGG	GTGCAACACC	8470								
ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540								
CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGCACCT	8610								
GGCCTATTTA	ACCATTTTAA	AACCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTT	8680								
AATTTCCCTT	TTACTCAGGA	GTTACCTTCC	TTTGATATTT	TCTGTAATTC	TTCTGTAGCT	GGGGATACAC	8750								
CGTCTCTTGA	CATATTACAA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	GGGCACTTGG	8820								
GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGGCG	8890								
AAGTGTGGAC	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960								
CTCCTACTCT	ACTGGGATTG	AGCCCTTCC	CTATCCCCC	CCAGGGGCG	AGGAGTTCT	CTCACTCCTG	9030								
TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACGTGCTGT	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100								
TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170								
GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240								
CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GSGGGTGGGT	GGGGTTTACC	9310								
ATGTTGGCCA	GGCTGGTCTC	GAACTTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCTCCT	AAAGTGCTGG	9380								
GATTACAGGT	GTGAGCCACC	ATGCCACGCT	CAGAATTAC	TCTGTTTAGA	AACTCTGGG	TCTGAGGTAG	9450								
GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAATT	TTTTTATTGT	TGTTAGAAC	9520								
CTCTTGATGT	TTTACTACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590								
ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCACAT	9660								
GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCT	TCTCTTCCCT	CTTTTAAAT	TGTTTCTTCT	9730								
ATGTTGGCTT	CTCTGCAGAG	AACCACTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800								
GGCCCTTTAG	CCTAGTGGCA	GAGACAATTC	ACAACACAG	CCCTTAAAAA	AGGCTTAGGG	ATCACTAAGG	9870								
GGATTTCTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940								
GGCCAGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAGGC	AATTTCTCTC	GGCAGTTTCT	10010								
GAAAGTAGTA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTCAAGTG	TTGCCGACCT	CAGCTACAGC	10080								
ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AACTTGAAGC	AACCCGGAGT	10150								
CTGGATTCTT	GCGAAGTCTT	CAGCTGCTCT	CGGTTGTGTC	CGGGGCCCA	GCTCTGGAGG	GGACCACTGG	10220								
CCGTGTGGCT	TCTACTGCTG	GGCTGGAGGT	GCGGCTCTCT	AGCTCTGCAG	TCCGAGGCTT	GGAGCCAGGT	10290								
GCCTGGAGCC	CGAGGCTGCC	CTCCACCTCT	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	CATCTGCCAG	10360								
ACAGAGTGCC	GGGGCCGAGG	GTCAAGGCGG	TGTGGCTGG	TGTGAGGCGC	CCGGTGCCTG	GCCAGCAGCA	10430								
GCGCTGGCT	CCATTTCCCA	CCCTTTCTCG	ACGGGACCGC	CCGGTGGGT	GATTAAACAGA	TTTGGGGTGG	10500								

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCGCTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCTCCGGGT	TGCTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCTTTC	ACGTCCGGCA	TTCGTGGTGC	10710
5	CCGGAGCCCG	ACGCCCCGCG	TCCGGACCTG	GAGGCAAGCC	TGGGTCTCCG	GATCAGGCCA	CGGGCCAAAG	10780
	GGTCGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCC	CGGGTTACCC	CACAGCCTAG	10850
	GGCGATTGCA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCGCGA	GGGGCGCGCG	10920
	GGCGGGGAAG	CGCGGCCAG	ACCCCCGGGT	CCGCCCGGAG	CAGCTGCGCT	GTCGGGGCCA	GGCGGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCACAG	ACCGCGCTCC	CCACGTGGCG	GAGGGACTGG	GGACCCGGCG	11060
10	ACCCGTCTGT	CCCTTTCACC	TTCCAGCTCC	GCCTCTCTCC	CGCGGACCCC	GCCCCGTCCC	GACCCCTCCC	11130
	GGGTCCCGGG	CCAGGCCCCC	TCCGGGCCCT	CCCAGCCCTT	CCCTTTCCTT	TCCGCGGGCC	CGCCCTCTCC	11200
	TGCGCGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GGATGCGCG	GCGCTCCCGG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCGC	GAGGTGCTGC	11340
	CGCTGGCCAG	GTTCGTGCGG	CGCTTGGGCG	CCCAGGGCTG	GCGGTGGTGC	CAGCGCGGGG	ACCCGGCGGG	11410
15	TTTCCGCGCG	CTGGTGGCCC	AGTGCCCTGT	GTGCGTGCCC	TGGGACGCAC	GGCCGCGCCC	CGCCGCCCCC	11480
	TGCTTCCGCC	AGGTGGGCCT	CCCCGGGGTG	GGCGTCCGGC	TGGGGTTAG	GGCGGCCGGG	GGGAACCCAG	11550
	CGACGCGCGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CCGCAAGTGT	CCTGCCTGAA	GGAGCTGGTG	11620
	GGCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGCGAAGA	ACGTGCTGGC	CTTCGGCTTC	GGCTGCTGG	11690
	ACGGGGCCCG	CGGGGGCCCC	CCCCAGGCCCT	TCACCACCAG	CGTGGCGAGC	TACCTGCCCA	ACACGGTGAC	11760
20	CGACGCACTG	CGGGGGAGCG	GGCGGTGGGG	GCTGCTGCTG	CGCCGCGTGG	GCGACGACGT	GCTGGTTTAC	11830
	CTGTGCGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGGCTCCCA	GCTGCGCCTA	CCAGGTGTGC	GGGCCGCCCG	11900
	ACGCGCCACT	CGGCGCTGCC	ACTCAGGCCCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGTCTGGG	11970
	ATGCGAAGCG	GCCTGGAACC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCGGGTGGC	12040
	AGGAGGCGCG	GGGGCAGTGC	CAGCCGAAGT	CTGCCGTTGC	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTTG	12110
25	AGCCGAGCGG	GACGCCCGTT	GGGCAAGCGG	CCTGGGCCCC	CCCGGGCAGG	ACGCGTGGAG	CGAGTGCACG	12180
	TGGTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCGGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTCTGGC	12250
	ACGCGCCACT	CCACCCATC	CGTGGGCCGC	GCCACACACG	CAGGCCCCCC	ATCCACATCG	CGGCCACAC	12320
	GTCCCTGGGA	CACGCCTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCCT	CAGGCGACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCCAGCCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGAG	GTTGCCCGCG	CTGCCCGAGC	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACCAACCG	CAGTGCCTCT	ACGGGGTGGT	12600
	CCTCAAGACG	CAGTGCCTCG	TGCGAGCTGC	GGTCACCCCA	GCAGCCGGTG	TCTGTGCCCC	GSAGAAAGCG	12670
	CAGGGCTCTG	TGGCGGCCCC	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GSTGCAGCTG	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCCTG	CCTGCCCGCG	CTGGTGGCCC	CAGGCTCTTG	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAAGTTA	TCTCCCTGGG	GAGAGTCCAG	12880
	AAGCTCTCGC	TGCAGGAGCT	GACGTGGAAG	ATGAGCGTGC	GGGACTGGCG	TGGCTGCGC	AGGAGCCACG	12950
	GTGAGCGAGT	GCTGGCCGTC	GAGGGGCCAG	TGAATGCAGT	AGGGGCTCAG	AAAGGGGGCG	AAAGGGGGCG	13020
	AGGCAGAGCC	CTGGTCTCTC	TGTCTCCATC	GTCACGTGGG	CACACGTGGC	TTTTCGCTCA	GGACGTGAG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGCATAA	ACTTACGAGG	TTACCTTTCA	13160
40	TGCTAACTCG	GACACGCGGT	TTCCAGGCCG	CGAGGCCAGA	GCACTGAACA	GAGGAGGCTG	GGCGCGGCTG	13230
	TGGAGCCGGG	TTGCCCGCAA	TGGGAGAGAAG	TGCTTGGAA	CACAGACGCT	CTGGCGAGGG	TGCTGTCAGG	13300
	TTACTATATA	GCTTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTTCTGCG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAAGTGCAC	GTGTCAGGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TGACCTGAGG	TCAGGAGTTT	GAGACAGCCG	TGACCAACAT	GGTGAACACC	TATCTGTACT	AAAAATACAA	13580
	AAATAGCTG	GGCATGGTGG	TGTGTGCCCT	TAATCCACGC	TACTTGGGAG	GCTGAGGCAG	GAGAATCACT	13650
	TGAACGAGGT	AGGCGGAGGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	CGACAACTGC	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GATAGAGGGA	GGGAGATAAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATGGGCTC	ACCTGCTGAG	GAAGGGACAG	TGTTTGTGGG	TGTTTGTGGG	ATGGTGTCTG	TGGGCCCTGC	13930
	CGTGTCCCGA	CCCTGTTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCACG	14000
	TGCTCCCGCA	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTACCCCCC	TCCCCACAAA	CTCCCCAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGCTC	TTTTTTTTCT	TTTTTTCTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACCTC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTG	CTTGAAATGC	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACCGG	AGCTGCCCTA	CACCTGCTGC	GGCTCAGGTG	GACCACGCGC	AGTCAGATAA	GCGTCATGCA	14350
	ACCCAGTTTT	GCTTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATT	CTGTTCTCTG	ACTTCAGATG	AGGTGACAAT	CTGCCCTTGG	CTTATGCAGG	GAGTGAGGCG	14490
	TGGTCCCGGG	GTGTCCTCTG	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTACGCTGTA	14560
60	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGGT	GAGTGAGGCG	14700
	TGGTCCCGGG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGCCCC	GGGTGTCCCT	GTACGCTGCA	14770
	GGGTGAGTGA	GGCGCGGTCC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGGC	CCAGGGGTGT	14840
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCAAC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCGGG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCGTG	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGCTCC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GGCGCGGTTG	15120
	CCCATTCGCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CAAAGCCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
	TCTTGGTCA	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGACTGCAGG	CTCTCGCCTC	CCGCGTGCCA	15260
70	GGCACTGCA	CCACAGCTTC	AGGTCGCTT	GCCTCTGTTG	GGCTTGGCTG	GCTCACACAG	TGCCCGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCAGCT	TGCTCATGCG	CGAGGCTGGA	CTCTGGGCTG	CCTGTGCTGT	15400
	CTGCCACGCT	TGCTGGAGA	CATCCAGAA	AGGGTCTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCAGAC	15470
	CCCTCAGCTT	GTCTGTGTTT	CTCCCAAGCT	GCCCTCTGCT	TTGGCCCCCT	TGGGTGGGTG	GCAACGCTTG	15540
	TACCTTATT	CTGGGACCTT	GCCGCTCAT	GCTTAGGCTG	GGCTCTGCTT	CCAGTCCGCC	CCTCAGATGG	15610
75	ATTGACGCTC	AGCCACAGGT	TGGAAGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGCT	GAGGGCCGCT	15680
	TTCTCCGCA	GCCTTCGTCA	GACTTCCCTC	TGCGGTCTTA	GTTTGAATT	TCACTGATT	ACCTCTGACG	15750
	GTCTATCTC	TCCATTGTAT	GCTTTTCTT	GTTTATTCT	TTCATCTCT	TTCTAGCTTC	TGATTTAGT	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCCCTTACC	TGCACCTGT	GTTTGTATGT	GAAGTAATCT	CAACATCAGC	15890

	CACCTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAA	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACTAT	GTTTATGTTT	AAGATATGTA	GAGTATCAAG	16100
5	ATACATAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATTT	CTAACTCAGT	TGTTAGTGGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTAATTTT	ACATCCTGTC	AATAGTGGGC	ATGCATGTTC	ACTATATCCA	GCTTATTAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAG	GAGGCCATAG	TCCCTCACCT	GGGGATGGG	16380
	TCTGTTCATT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
10	TTATCTTCCCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTGTC	TGTGTCTGTT	TTCTGCCCTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCGCCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACITT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCCTTA	CACCTGGCTA	ATTTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
	TGCTGTGTTG	CCCAGGCTGG	TCTCAAACCT	TTGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
15	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAAATTT	CAACACTTTT	ATATTCTTAT	AGTGTGGGTA	16940
	TGTCCCTGTA	ACAGCATGTA	GGTGAATTTT	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CTCTGTTCCC	TTGTTTCTCA	CCACCTCTTG	GGTGTGCCATG	TGCGTTTCTCT	GCCGAGTGTG	TGTTGATCCT	17150
20	CTCTGTTCCC	CTGTTTCTCA	CCACCTCTTG	GGTGTGCCATG	TGCGTTTCTCT	GCCGAGTGTG	TGTTGATCCT	17220
	ATTGTGCTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTACCCCA	GGCTGGAGTG	TAATGGCACA	17290
	ATCTCGGCTC	ATGCAACCTT	CTGCTCCTCT	GGTTCAGCA	GTCTCATTC	CTCAACCTCA	TGAGTAGTGC	17360
	GGATTACAGG	CGCCACCCAC	CACGCTGGGC	TAATTTTTGT	ATTTTTAGTA	GAGATAGGCT	TTCAACCATGT	17430
	TGGCCAGGCT	GGTCTCAAAC	TCCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TTGATCTTTT	AAAATGAAGT	CTGAAACATT	GCTACCCCTG	17570
	TCTGAGCAA	TAAGACCCCT	AGTGATTTTT	AGCTCTGGCC	ACCCCCAGC	CTGTGTGCTG	TTTTCCCTGC	17640
	TGACTTAGTT	CTATCTCAGG	CACTTTGACA	CCCCCACAAG	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTTCCT	GTAGCTTTTG	CCCCGCCCTG	CTTTTCTCTC	TTTGTTCCTC	GTCTGTCTTC	TGCTCAGGCT	17780
	CGCCGCTCTG	GGGTCCCTCT	CCTTGTCTCT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
30	CTTTACTCTG	GCTGGCCTCC	ATGGCATCTA	GCGACGCTCG	GGGACCTCTG	CTTATGATGC	ACAGATGAAG	17920
	ATGTGGAGAC	TCACGAGGAG	GGCGGTCTATC	TTGGCCCGTG	AGTGTCTGGA	GCACCCAGTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGGCCTGGGT	TCAGCCTGGA	AAACCCAGG	CATGTCGGGG	18060
	TCTGTGGCTG	CCGCGGTGTC	GAGTTTGAAG	TCGCGCAAAAC	CTGCGGTGTG	GCGCCAGCTC	TGACGGTGTCT	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCTCTT	TCTGCTTGGG	AACCCAGGACA	AAGGATGAGG	CTCCGAGCCG	18200
35	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTAAAT	TCTAGGCTGG	GCGCGGTGGC	18270
	TCACGCTCTG	AATCCAGCA	CTTTGGGAGG	CCAAGGCGGG	TGGATCACGA	GGTCAGGAGG	TGAGACCAT	18340
	CCTGGCCAAC	ATGATGAAC	CCCATCTGTA	CTAAAAACAC	AAAAATTAGC	TGGGCGTGGT	GGCGGGTGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	TTTGCAGTGA	18480
	CGCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAATA	18550
40	AAAAAATAAA	AATCTAGTA	GGCACAATAA	AAAAAGTAAA	AAGAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCAGCATG	TCCACACCTC	ATCATTTTAA	GGTGTATTG	GTGGGAGCAT	CACTCACAGG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGGTCCC	GTGCGTGCC	ATCTCGGCTC	18760
	GGACCTGGCT	GGCTTCCCAT	GGCCATGGCT	TTTGTACCAG	ATGGTGCAGG	TCCGGGATGA	GGTCGGCAGG	18830
45	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGCACG	TCTGGGATGA	GGTCGCCAGG	CCCTGCTGTG	18900
	AGCTGCGCTA	GTGGTGTCTG	GATGGTGCAG	GTCAGGGGTG	AGGTCTCCAG	GCCCTCGGTG	AGCTGGGCTG	18970
	ATGGAGTCCG	GATGATGCAG	GTCGGGGGTG	AGGTGCGCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	TGTAATCCCA	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGG	AGGTCTGGGG	19180
50	TGAGGTCAAC	AGGCCCTGCG	GTCAGCTGGG	TGTGCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTCCG	19250
	CAGACGGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTGG	19320
	CCAGGCCCTG	CTGTGAGTTG	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAAGTCT	GGGGTGAAGG	TCCCGAGGCC	CCTGCTTGTG	19460
55	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCTGGAGTG	AGGTGCGCAG	GCCCTCGGTG	AGCTGGATGT	19530
	CGAGTGTCCA	GATGGTGCAG	GTCGGGGGTG	AGGTGCGCAG	ACCCTGCGGT	GAGCTGGATG	TGCGGTGTCT	19600
	GGATGGTGCA	GGTCTGGAGT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19670
	GGTCCGGGGT	GAGGTGCGCA	GACCCTGCTG	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TAGAGTCCGC	AGACCCTGCT	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTGAGGG	GTGAGGTCTC	19810
60	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTGC	19880
	TGTGAACCTG	ATGTGCGCGG	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTGC	CCAGGCCCTC	GGTGAGCTGG	19950
	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTGC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	20020
	GTCTGGATGG	TGCAGGTCTG	GGTGTGGTGC	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGTCCG	20160
65	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGGTTCT	GGGTGAGGTT	20230
	CACCAAGGCC	TGCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAAGTC	CGGGGTGAGT	TGCGCAAGGC	20300
	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGCT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
	TGGTGGGCTG	GATGTGCCGT	GTCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	20440
	GATGTGCCGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGTGGT	20510
70	GTCCGGATGG	TGCAGGTCCG	GCTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGCTGTGAGT	20580
	GTGCAAGTCC	GGGGTGAAGT	AGCCAAAGGCC	TTCCGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGCAAGTC	20650
	CGGGGTGAGG	TGCGCAAGGC	CTGCGGTTAG	CTGATATGTC	GGTGTCCGGA	TGGTGCAGGT	CCGGGGTGAG	20720
	GTCACCAAGC	CCTGCGGTTA	GCTGGATGTG	CGGTGTCTGG	ATGGTGCAGG	TCCGGGGTGA	GGTGCAGG	20790
	CCCTGTCTGT	AGCTGGATGT	GCTGTATCCG	GATGGTGCAG	GTCGGGGGTG	AGGTGCGCAG	GCCCTGCAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGA	GAGTGTGCGG	GAGGTGCGCA	GGCCCTGCGG	TGAGTGGAGT	20930
75	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTGACCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
	CGGATGGTGC	AGGTCTGGGG	TGAGGTGCGG	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGATGGTGT	21070
	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTGC	GGTGAGCTGG	ATGTGCTGTA	TCCGGATGCT	GCAGGTCTGG	21140
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAAT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGCTGGATG	GTGCAAGGTCC	GGGGTGAGTT	GCCAGGCCCT	21280

	TGCGGTGAGC	TGGATGTGTG	GTGTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCGGT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAAGCTG	GATGTGCGGT	21490
	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCTG	21560
5	TTTAAGGGGT	TGGCTGTGTT	CCGGCCCGAG	AGCACCGTCT	GCGTGAGGAG	ATCCTGGCCA	AGTTCCTGCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTCCGAGCT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAACAGGC	TCCTTTTCTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAAAG	CATTGGAATC	AGGTACTGTA	21770
	TCCCCACGCC	AGGCCCTCTG	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TCAGCATGCG	CCTGTCTCCA	21840
	CTTGCCCTGTG	CTTCCCTGGC	TGTGCAGCTC	TGGGCTGGGA	GCCAGGGGCC	CCGTACAGAG	CCTGGTCCAA	21910
10	GTGGATTCTG	TAGCAAGGCTC	TGACTGCCTG	GAGCTCACGT	TCTCTTACTT	GTAATAATCAG	GAGTTTGTGC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTTAAA	TTAGATGGAA	ACACTACCAC	TAGCCTCCTT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTTGCCCAG	GCTGGAGTGC	AGTGGCATAA	TCTTGGCTCA	CTGCAACCTC	CACCTCCTGG	GTTTAAGCGA	22190
	TTACACAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCTTGGCT	AATTTTGTGA	22260
15	CTTTAGGAG	AGACGGGGT	TCACCATGTT	GGCACGGCTG	GCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGGC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCCGAT	TCTCTTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGATTTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAGGGAGC	ACCTGTGCAG	GGAGACCTTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
	TCTAGGTGGC	TGCATTGAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCCG	CTGATGAGAG	TGTGAGATTG	22610
20	TGACAGATTG	AAGCTGGATT	TGCATCAGTG	AGGGACGGGA	GCGCTGGTCT	GGGAGATGCC	AGCCTGGCTG	22680
	AGCCCCAGGC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTCAGAAGAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
	CGTGGTCTG	GGGCCATTTC	CTTGATCTCG	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AAGTTAATAC	22960
25	ACAATGACC	TTACTTAGAC	TTTACACGTA	TTTAATGGTG	TGCCACCCAA	CATGGTCATT	TGACCATGAT	23030
	TTTGGAAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCCAAGA	TGCTCCTTGT	23100
	CACACTCTGG	ACTGTTGTTT	TGCTTGGGGG	GCTTGGGAGG	CCCCCTCTCC	CTGGACAGGG	TACCGTGCTT	23170
	TTTCTACTCT	GCTGGGCCGTG	CGGCTCTGCG	TCAGGGCACCC	AGCTCCGGAG	CACCCGCGGC	CCAGTGTCTC	23240
	ACGGAGTGCC	AGGCTGTACG	CCACAGATGC	CCAGGTCCAG	GTGTGGCCCG	TCCAGCCCCC	GTGCCCCCAT	23310
30	GGGTGGTTTT	GGGGGAAAG	GCCAAAGGCA	GAGGTGTACG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGCC	TGAGCAGCCT	CTCCCGCCT	CTCCATCTGA	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGGTCC	TGCCCTGGGG	CAGCCTTGGG	CTACCCCACT	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGCAT	GGGTTACAGT	GGCCCCAGAT	CAGCCTTGGG	ACCAGGCTCC	CTGGTGCTGA	TGGTGGGACA	23590
	GTCACCTTGG	GGGTTGACCG	CCGGAAGTGG	CGTCCCAAGG	GTTGACTATA	GGACCAAGTG	TCCAGGTGCC	23660
35	CTGCAAGTGG	AGGGGCTCTC	AGAGGCGTCT	GCTTGGCATG	GGTGGACGTG	GCCCCGGGCA	TGGCCTTCAG	23730
	CGTGCTCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCGG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGTCTGGCTG	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAAAGGGGTG	23870
	CAGCTGCGGG	AGCTGTCCGA	AGCAGAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCCGCTTG	CTGACGTCCA	23940
	GACTCCGCTT	TCTCCCAAG	CCTGACGGGC	TGCGGCCGAT	TGTGAACATG	GACTACGTCG	TGGGAGCCAG	24010
40	AACGTTCCCG	CAGAAAAAGA	GGGTGGCTGT	GCTTTGGTTT	AACTTCTTTT	TTAAACAGAA	GTGCGTTTGA	24080
	GCCCCACATT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTTGT	GCGCACAGTG	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTGTAGGG	GGAGCTCCTG	GGGCAAGGAC	AGGCTCTGAG	GACCACAGAA	AGCAGCCGGG	CCAGGGCTTG	24290
	GATGCAGCAC	GGCCCGAGGT	CCTGGATCCG	TGTCTGTCTG	TGGTGGCGAG	CTTCCGTGCG	CTTCCGCTTA	24360
45	CGGGGCCCGG	GGACAGGGCC	ACGACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAC	CTTGCCCTAC	24430
	GGCTTCCTGCA	CCCCACCCCT	GTGGCTCTGG	TGGGCTCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAAAACAG	24570
	GGTCTAGAGA	AGCTGGGAGG	GGTTCTAGGT	CCCGGCTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
	TCTCCCTTGG	GTCCCTATGG	TGGGCTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
50	CCCCGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTACAGCT	GCTCAACTAC	GAGCGGGCGC	24780
	GGCGCCCCGG	CCTCTTGGGC	GCCTCTGTGC	TGGGCTTGGA	CGATATCCAC	AGGGCCTGGC	GCACCTTCGT	24850
	GCTGCGGTGT	CGGGCCCAAG	ACCCGCCGCC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGCCGG	GGACCCCGCT	24920
	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
	GGGTGGGCGG	CAGGAGTGTG	AGGTGACCCCT	GTCACCTTGT	AGGACACACC	TGGACCTTAG	GGTGAGGGCC	25060
55	TTACGCTTTT	CCTGCAGCAC	ATGGGGCCGA	CTGTGCACCC	TGACTGCCCC	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTACAGAA	AGGAACCGCA	ACGGCTCAGC	CACCAAGGCC	CGGTGCCCTT	25200
	CACCCAGATC	CTGAGCCAGG	GGTCTCTGTG	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
	GGGTGTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTG	25340
	ATGTCTGAGT	TCTTGGCTGG	CCACTGTCTG	TCTCTCGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCG	25410
60	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTCTATC	GCCAGCATCA	TCAAACCCCA	GAACACGTAC	25480
	TGCGTGGCGT	GGTATGCCGT	GGTCCAGAAG	CCCGCCCATG	GGCACGTCCG	CAAGGCCCTT	AAGAGCCACG	25550
	TAAGGTTTAC	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCAGTC	25620
	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGTGGCA	25690
	CGTGTGTGTC	GTGGTGCATG	TATCTGTGGC	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
65	CCATGGGTGT	TGTGCTCTGT	GTGTGCATGT	GTGTGTGTCT	GTGACACGTG	CATGTTTCTG	CTGTGTGTGT	25830
	CATGTCTGTG	ATGTGCCTAT	TGTGGTGTG	TGTGTGCATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCTCTTGGC	CTTACTCTCT	CCTCCTCCAG	GCATGGTCCG	CACCATTTGT	CTCAGCTCTT	25970
	CGGTGTCTGG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TCTAGCATG	GGTGGCCCTG	TCCTGTGACA	26040
	GGGTGTGGCC	TTGGAGACTG	TAAGCCAGGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTTGGAC	26110
70	CCCTGGACAC	CCCAGGACCC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATA	TAGGAAGGCT	GATTACAGCC	26180
	TCGCTGCCCG	GGACACACTC	CTCCAGAGCC	GGCCGGGGCC	CTTGGGGCTC	GGCAGGGGCT	AAAGGGGCCC	26250
	TGGGCTTGGG	TTCCCAACCA	GTGGTCTATG	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGGCC	26320
	GGGTGCAGAG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCCAAGGA	26390
75	CCTCTTTCTC	TGACTTCTTG	AGCT					26414

Contig 2:

	TGTGGGATTG	GTTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTCAAG	GCGAGCTTTC	TTCTCTGTAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAT	140
5	ATCTTCCCTTT	GAACATATGGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTGGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACCTCTGGG	TCGTGTGGTG	CCTGCTGTGG	TGTGTGGCCG	GTGGGCAGGG	CTCCAGGGCC	280
	TCCTTGTGTT	CATTGGCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTCTGG	AATTCCTCTG	CGAGTGGGAG	350
	GCCTTCTTTC	TTTTCTTTTT	TCTTTCTTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCCTG	AGTTCAAGCA	ATTCTCTTGC	490
10	CTCAGCTCC	CAAGTAGCTG	GAATTATAGG	CGCCCAACCAC	CATGCTGACT	AATTTTTGTA	ATTTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCCTC	CCACCTCGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCCGGCC	CGGCCGAGAC	TCGCTTCTCT	CAGCTTCCGT	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGTGAC	AACCTCCGTT	TTCTTCTCTC	AGGTCTCGCT	770
	AGGGGTCTTT	CCATTTTCATG	ACTCTCTTCA	CAGAAGAGTT	TCACGTGTGC	TGATTTCCCG	GCTGTTTCCT	840
15	GGTAATTGG	TGTCTGCTGT	TTATCGATGG	CCTCCTTCCA	TTTCTTTTAG	GCTTTGTTTA	TTGTTGTTTT	910
	TCCGGCTCCT	TGAAGGAAAA	GTTTCGATTA	TGGATGTTTG	AACCTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCCCTTTTC	CCTCTAAAGC	AGGGATCCCG	AGGCCCTTGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTT	1050
	AGGAACCCGG	GCGACAGCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCACGGT	TCCCGCCTGA	GCCCCGCCCC	1120
	TCTCAGATCA	GCAGTGGCAT	GCGGTGCTCA	GAGGCGCACA	CACCTTACTG	AGAACTGTGC	GTGAGAGGGG	1190
20	TCTAGATTCT	GTGCTCCTTA	TGGGAATCTA	ATGCCGTGATG	ATCTGAGGTG	GAACCGTTTG	CTCCCAAAAC	1260
	CATCCCCCTC	CCCCTGCTG	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACAATGG	1330
	TTGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTCACTGTTG	ATATATTGGC	TTTTCTGTGT	1400
	TGAGTCCAGA	ATAATTACGG	ATTCTGTGTA	TGCTTCCCG	CGACCTCAGA	CCCATGGGCT	ATTTGTGGGC	1470
	GTGTGTGCTG	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCAT	GTACCTTCTC	GTTACTGCCT	TCCAGGTTGG	1540
25	TTCTCAGGCT	TGAATCGTAC	TCGATGTGCT	TTTAGCCAC	GGCCCTGCCG	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGCTGA	AGCACAGAGT	CACCGTGCCG	GTCTTTTGAT	GCCTCACAG	CTCGAGGCTT	CCTGTGCTCG	1680
	TGTAGTGTG	TGTCACTGCT	CTGCTCACAT	CCTGCTCTGG	GGACGACAGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCTGGGGG	AGTCTGCAGA	ATAGGAGGTG	GGGGTGCCGG	TCTCTCTCCC	CGGTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTGCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGAGAGGGC	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCCGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTTGG	TCACAGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCG	TGGGGTCTGA	2030
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCG	TGGGGTCTGA	2100
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCG	TGGGGTCTGA	2170
35	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGTT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2240
	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGTT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2310
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCG	TGGGGTCTGA	2380
	TGTGTGTTGA	CTGTGGATGG	CGGTCTGTGG	GTCTGTGTG	GTGACTGTGG	ATGGCGGTCG	TGGGGTCTGA	2450
	TGTGTGTTGA	CTGTGGATGG	TGATCGGTCA	CAGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	CGGGTCTGAG	2520
40	GGTCTGATGT	GTGGTGAAGT	TGGATGGTGA	TCGGTCAACG	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	2590
	GTCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTTG	GTCCCGGGGG	TCTGATGTGT	GGTGAAGTGT	2660
	GATCGGATC	GGTACACAGG	GTCTGATGTG	TGGTGAAGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGTT	2730
	GACTGTGGA	GGCGGTCTGT	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	GTCTGGGGT	CTGATGTGTT	2800
	GACTGTGGA	GGCGGTCTGT	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	GTCTGGGGT	CTGATGTGTT	2870
45	GACTGTGGA	GGCGGTCTGT	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	GTCTGGGGT	CTGATGTGTT	2940
	GTGGTGAAGT	TGGATGGCAG	TCGTGGGGT	TGATGTGTGG	TGACTGTGGA	TGGCGGTCTG	GGGGTCTGAT	3010
	GTGTGGTGA	TGTGGATGGC	GGTCTGGGG	TCTGATGTGT	GGTGAAGTGT	GATGGCGGTC	GTGGGGTCTG	3080
	ATGTGTGGTG	ACTGTGGATG	GGGTCTGTGG	GGTCTGATGT	GGTGAAGTGT	GATGGCGGTC	GTGGGGTCTG	3150
	ATGTGTGGTG	ACTGTGGATG	GGGTCTGTGG	GGTCTGATGT	GGTGAAGTGT	GATGGCGGTC	GTGGGGTCTG	3220
50	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	GTCTGGGGT	CTGATGTGTT	GACTGTGGAT	GGCGGTCTGT	3290
	GGGTCTGATG	TGTGGTGAAGT	GTGGATGGCG	GTCTGGGGT	CTGATGTGTT	GACTGTGGAT	GGCGGTCTGT	3360
	GTCAACAGGG	TCTGATGTGT	GGTGAAGTGT	GATGGCGGTC	GTGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	3430
	GGCGTCTGTG	GGTCTGATGT	GTGGTGAAGT	TGGATGGCAG	TCGTGGGGT	TGATGTGTGG	TGACTGTGGA	3500
	TGGCGGTCTG	GGGGTCTGAT	GTGGTGAAGT	TGGATGGCAG	TCGTGGGGT	TGATGTGTGG	TGACTGTGGA	3570
55	GCAGGTGGAG	TCCAGGTGT	GTCTGTAGCT	ACTTTGCGTC	CTCGGCCCCC	CGGCCCCCGT	TTCCCAACA	3640
	GAAGCTTCCC	AGGCGCTCTC	TGGGCTTTC	CCCGCCATCG	GGCTTGGCCG	CAGGTCCACA	CGTCTGATC	3710
	GGAGAAACA	AGTGCCACG	TCTGGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCCCGC	3780
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCGT	ACATGCGACA	GTCTGGGCTC	CACCTGCAGG	AGACAGGCC	3850
	GCTGAGGGAT	GGCGTCTGCA	TCGAGCAGGT	CTGGGCACTG	CCCTGCAGGG	TTGGGCACGG	ACTCCAGCA	3920
	GTGGGTCTCT	CCCTGGGCAA	TCACTGGGCT	CATGACCGGA	CAGACTGTTG	GCCCTGGGGG	GCACTGGGGG	3990
60	GAATGAGCTG	TGATGGGGG	ATGATGAGCT	GTGTGCCTTG	GCGAAATCTG	AGCTGGGCCA	TGCCAGGCTG	4060
	CGACAGGCTG	TGCATTACAG	CACCTGCTCA	CGTTTGACTG	CGCGGCTCTC	CTCCAGTTCC	GCACTGCCCT	4130
	TGTTTATGAT	TGTGTAATG	TCTTCTCTGC	CAGTTTGTAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCCCTCC	4200
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCCTGCCACG	CTGGCCCTCT	AGTGTGGGTT	CTGAGGCCAA	4270
65	AGGAAACGTG	TCCCTTCTCT	TAGGAGGACG	GGCGGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCTCTCAG	4340
	TGCTGGGCTC	GTCCACGTGG	CCCTGTGGCC	CTTTGACAGT	GTGGTCTGTC	CACGTGGCCC	TGTGGCTCTT	4410
	TGCAATGCTC	TGTTAGCACT	TGCTCGGCTC	TAGGGGACAG	TCGTGTCCAC	CGCATGAGGC	TCAGAGACCT	4480
	CTGGGCGAAT	TTCCTTGGCT	CCCAGGGTGG	GGGTGGAGGT	GGCCTGGGCT	GCTGGGACCC	AGACCTGTG	4550
	CCCGGAGCTG	GGGCGCAAC	TCTTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGCTGTG	TGGGTGTGAG	4620
70	CCAGCTGGA	CCCACAGGTG	GCCAGAGGA	GACGTTCTGT	GTACACACT	CTGCCAAGC	CCATGTGTGT	4690
	CTGCAGAGAC	TGGGCCCCGC	CAGCCACAGA	TGGCCCTGCA	TTCCAGCCCA	GCCCGCAGT	TCATCACAAA	4760
	CACAGACCCC	AAAAGGGACG	GAGGGTCTTG	GCCACGTGGT	CCTGCCTGTC	TCAGCACCCA	CCGGCTCACT	4830
	CCCATGTGTC	TCCCGTCTGC	TTTCGACAG	CTCTCCCTTG	AATGAGGCCA	GCAGTGGGCT	CTTCGACGTC	4900
	TTCTACGGT	TCATGTGCCA	CCACGCCGTG	GCAATGAGT	CAGGTGGCCA	GTTGTGTTAT	GTTGTGTTAT	4970
	CCCTGCGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCTT	GCCCTTCCC	CACTGNCCTT	5040

	CTGCCCGGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCCG	5110
	AGGCCCGGAA	ACATGGCTCG	GCTTGCCGCA	GGCCGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
	AAGCGGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGCC	5250
5	GAGCGTTTGA	GCCTGCAGCT	TGTGAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGCTC	TCACACGCTT	5320
	GTATCTCTCT	CTCCCGATAC	AAAAGGATTT	TATCCGATTC	TCATTCCTGT	CCCTGTCTGT	TGACCCCGCC	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAAGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCGG	GGGGCTCTGT	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTACAC	CAGCCACTGG	5530
	GTGAGCCACA	CTCAGGGTGG	TAGAGCCACA	GTGCCTGGTG	CCACATCACG	TCCTCTGGAT	TTTAAGTAAA	5600
	ACCACACACC	TCCCGGCGAG	CATCTGCTGT	CGACCCTGTG	TGTGCCTGGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTGCTGCA	CACCTCAAGT	CATCAGCAAG	GTATCCGCA	GTCAGGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCAATTAT	AGTAATACAC	TAATGGTATC	AGCAATTATA	ATATTTTATA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGCCC	GAGGGACACG	5950
15	TGTGCACATG	TGTGTAAGCG	GCCCCCAGGC	CCACAGAAAT	CGCTGACAAA	GTCACCTCCC	CAGAGAAGCC	6020
	ACCACGGGCC	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	GTCTCTATGC	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	CTGAATAAAA	6230
	ACGTTCTTAA	AACCTGTTGC	CCCAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GCCGGACTCC	TAGAGTTGGT	GGGTGTGCTT	CTGTGCAAAA	6370
	AGTGCAGTCC	CTTTGCCCAT	CACGTGTGTA	TCTGCACCAG	CAAGGAARAG	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACTGTTG	TCTGCCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCTGCCT	CAGCCTCCCG	AGCAGCTGAG	ATTACAGGCA	6580
	CCCAACCCCT	CGCCCTGGCT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTT	TTGCCATGTT	GGCCAGGCTG	6650
	GTCTCGAATC	CTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
20	CCATCGAACC	CAGCCGGAAC	GCCTCTTTT	AAGGTGACCA	CCTATAGCGC	TTCCCGAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTT	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACGGGG	GCTATTCTGC	TCTCACTGTT	6930
	TGCTGAAAA	CGCACCCCTG	GCATCCCTGT	TTGGAGAGTT	TCTGCTTCTC	GTGGTCAATG	CTGAAAGTAG	7000
	GGGCAAGGTT	GTATCCGTTG	GCGCGCAGCG	GCTACATGTA	GGGTCAATGAG	CTTTTCACCG	TGGACAAATT	7070
30	CCTTGAAAAA	AAAAAAGGA	GTCCGGTTAA	GCATTCAATC	CGGGTCAAGT	GTCTGTTCT	GTGAATAAAC	7140
	TCTAAGATTT	AAGAAACCTT	AATGAAAGAA	AACCTTGATG	ATTGAGAGCA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTACGCGGC	CCCAGTGCAT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGT	CAGAGGTCTC	7280
	ATCTGGTATG	TTTCTGAGGT	GTTTGCCGGC	TGAATGGTAG	ACGTGTCTGT	TGTGTGTATG	AGGTTCTGTG	7350
	TCTGTGTGTG	GCTCGGTTTG	AGTGTAACGA	TGTCCAGCAC	ATGCCCTGCC	GCTCTCTCAC	CTGTGTCTTC	7420
35	CGGCCCGGCG	TCTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTGAGC	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGCGGGGA	TTCGGCGGGA	CGGGTGAGGC	CTCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTTGATTGCG	TTTTGATGCA	TTCAGTGTTA	ATATTCTGGG	TGCTCTGGAG	7630
	ACCATGCTAT	CTCTGTCTTG	AGGAACAGCA	CAAGGTTGCA	GCCCTTCTCT	GGTATGAAGC	CGCACGGGAG	7700
	GGGTGACACA	GGCTGAGGAC	TGCGGGCTCC	ACGCAAGGCTC	TGTCCAGCGG	CCATGTCCAG	AGGCCTCAGG	7770
40	GCTCGAAGAG	CGGGAGGGCC	GCTGCGCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCAACCCAG	7840
	CTTCTGTCTC	GTCAACCCAG	TTCCGTTAGG	GTCTTGGGG	AGATGGGGCT	GGTGCAGCCT	GAGGCCCCAC	7910
	ATCTCCGAGC	AGGCCCTCGA	CAGGTGGCCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCA	TCCCACTTGC	7980
	ATGGGGTCTA	CACCCAAAGGA	CGCACACACC	TAAATATCGT	GCCAAACCTAA	TGTGGTTCAA	CTCAGCTGGC	8050
45	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGCAG	8120
	GTTAGTTACA	TATGTATACA	TGTGCGATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGTAT	8190
	TATCTCTCTA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCCCACC	8260
	CTGTGTCCAA	GTGTTTCTCAT	TGTTCACTTC	CCACCTGTGA	GTGAGAATAT	GTGGTGTGTT	GTTTTCTTTC	8330
	CTTGCAATAG	TTTGTCTAGA	GTGATGTGTT	CCAGCTTCTG	CCATGTCCCT	ACAAAGGACA	TGAACCTATC	8400
50	CTTTTTTATG	ACTGCATAGT	ATTCCTGGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTTGG	GTTGGTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCTTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAATGGGTAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAAATCACCA	CACCTGTCTC	CACAATGGTT	GACTAGATTT	AACTCCCAAC	8680
	CAACAGTGTA	AAAGTGTCT	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
55	CACCTCCGGT	TTTTCTGTG	CATCTTTTGA	AACTTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCTAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTGTACAA	8960
	CACAACCTGC	TCTGGGATT	GGAGGAAAGT	GTCCCTGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGCGC	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCGAG	9100
	CTTCTGTG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
60	ACGTGGCTCA	GAGGGGGCGA	GGTTCCAGC	CCACGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCCTAAGAG	9240
	TCTGAGAAAT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTG	9380
	AAAGAAATTT	AAGTTTTTCA	TTTAACCGCT	TTGGAGAAATG	TTACTTTTAT	TATGGCTGTG	TAAATTGTTT	9450
65	GACATTGAGT	CCCTCGTAGA	CAGATCACTAC	GTAAAAAGTG	TAAAGTTAAT	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCTCTGCT	TTGGTGGATG	ATTTCTGTGT	GGTGACACCT	CACTCAACCC	ACGCGAAAAC	9590
	CTTCTCTAGG	TGAGGCCCCG	GCCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTTGCC	ACCGCAGCGT	TGTCTCTGCC	AAGTCTCTCT	TCTCTGCCGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCGCTTGG	CCGTGACACC	AGGCGTGGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
	ACCGTGAGCG	CCCTGTGCTT	GCAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTAGCTGGG	9870
70	CTCCTGCTGC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCA	GGAGCTGTGC	9940
	ACAGGGGCTG	CAGGGCCGAG	GCGGCACGCT	CTCCCCAGG	GTGCACCTGA	GCTTGGGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGGCCACA	GCGTTCGCTG	CGGTACAGTT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TGCTGAGTG	CTGCTGTCTT	GAACCCAGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
	TTTGTGAATC	AAACTAAAT	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTCCA	ATGTGGTCCC	10220
75	CCTCAAGGCG	GCCCCACAGA	GCCGGTGGGC	TTGTTTTAAA	GTGCGATTG	ACGAGGGACG	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	GGCCGCCAGG	GGTGGTTTCA	GGTGTCTTGC	TGGGCTGTGT	TGGCGTGTG	10360
	TTGTGAAAC	CCATTGGAC	CCGCCCTCCA	AGTCCACCTT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430

	TG	GGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCCGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AG	ATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AG	AGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTCTT	GAGGGTGCTG	GCCAGGGAGG	TGGCTCAGAG	10640
	TG	TATGTTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTGGCG	AGCCATGTAA	CAGGAAGGGG	10710
5	TG	GCCACAGG	GAGCTGGGAA	TGCACCAAGG	GAGCTGCCGA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GA	AGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCCGA	GGAAAGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAG	GCTACCG	GGCAGAGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCCA	GGGAACCTCC	10920
	TT	GACGTGAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAACT	10990
	CAG	AACCCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCCTT	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTCG	11060
10	TT	GAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTGTAGTCTG	CCCCGGACCA	11130
	CAG	ATGAGTCT	TATAACGGGA	TTGTGGTGT	GGCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GG	GCGTGCAC	CTCCCATCTG	AGTCTTGCT	GTCCCGGCTC	CAGGCCAGGT	TCTTGCATGC	TCACCTACCT	11270
	GT	CCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CT	GCCAGGCT	CAGCACCCCTG	CTCCAAATCA	CCACTTCTCT	GGGGTTTTTC	AAAGCATTTA	ACAAGGGTGT	11410
15	CAG	ATGACCT	CCTGGGTGAC	GGCCCCGCT	CCTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCTGGTCT	11480
	CG	AGGTGCTC	CTGAGTATGG	CTGCGTGGTG	AACTTGCGGA	AGACAGTGGT	GAACTTCCCT	GTAGAAGACG	11550
	AG	GCCCTGAG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCTTATTC	CCCTGGTGCG	GCCTGTGCT	11620
	GG	ATACCCCG	ACCCCTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCCGGAAG	TGGAGCCCTGT	11690
	GCC	CGGCTG	GGCAGGTGCT	GCTGCAGGGC	CGTTGCGTCC	ACCTCTGCTT	CCGTGTGGGG	CAGGCGACTG	11760
20	CCA	ATGCCCA	AGGGTCAGAG	GGCACAGGGT	GGCCCTCGTC	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CT	GTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAAAGGAA	ATGGTGCACT	11900
	AG	ACCTGGGT	GCACGTAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGCTG	11970
	GG	CGTGAGTC	TCTCAAACCC	GAACACAGGG	GGCCTGCTGG	GCATGAGTCC	CTCTGAACCC	GAGACCCCTG	12040
	GG	CCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
25	GT	GAGCCCCA	CACCTCAAGG	CTCATCCACA	CCTACACAGG	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TC	AGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTCCTC	CAGAGCCCGA	12250
	GAG	CTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAATGAATT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TT	TTTTTTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCCAGC	12390
	ACT	TTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAACC	AACATAGTGA	12460
30	AA	TTCCATT	CTACTTAAAA	AATACAAAAA	TTAGCCCTGG	CTGGTGGCAC	ACGCCCTGAG	TCCCCTGTAT	12530
	GCG	GGAGGCT	GAGGCAGGAG	AATCATTTGA	ACCCAGGAGG	CAGAGGTTCG	AGTGAGCCGA	GATCACACCA	12600
	CT	GCACTCCA	GCCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAATAAA	AAAAAGTATC	AGCATTTCCA	12670
	AA	CCATAGTC	GACAGGTGTT	TTTTTATTCT	GTCCCTTCGAT	AATATTTACT	GGTGCTGTGC	TAGAGCCCTG	12740
35	AA	CTGGGGT	GCCTTCTCT	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTTAAA	12810
	CC	AGAGGTTT	AAACTGGGGT	CCTGTCTGTC	TGAGTTAAAC	GTCCAGATCT	GGACTTTGCC	TCTTCCGAGA	12880
	AT	GCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	CG	ACAGCCCG	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCCAAG	ATGTCCCTGT	13020
	TG	CAGCTCCC	TCCCCACAA	GAGTCCCGTC	TCCTGTGCTC	CCACAGTCCC	CTGCTTCCCT	CTCACAGCTC	13090
	TAC	CTGTGCTC	TGGCCTCCAC	TGGCTTTGTC	TGCTATGATT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
40	GC	CTCTCCCA	GGCACCCTCG	CAGTGGTGGC	CATACCAGTC	AGCTGTGAAC	TGTCCACTGC	TTATTTTGCT	13230
	CCCC	ATGAAA	TGTATTTTT	AGGACAGGCA	CCCCGTGGTC	CAGCCTCTGG	CACAGCATCA	GTGAATGTTA	13300
	TT	GAGAGACA	AAGGACAGAC	AAACAAATCA	GGAAATGGG	TTCTCTTAA	ACACATTGCA	AAGCCACAGA	13370
	GG	GATAGTGA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAATATTC	TGTGCTCCCA	13440
45	AA	GGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAGCTCA	GCAGTGGAGG	CAGTGGTTCG	13510
	CC	ATACTCAG	GGTGAATCA	CATCCTCTGT	GTCTGAAAGT	TACAGCAGAG	GCTTGAAGGG	CATCTGGGAG	13580
	AA	GAACACAG	GCAAAATGAT	TAAGAAAAGT	GAAAGAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	AT	TTTTAGTCT	CCCAAACAC	AGCTCAGATG	GTAGAATGTG	GTCAGAACTG	ATGGACAGAA	CAATAGAACA	13720
	AA	ACGAGAGC	CCTATCTCTC	AGAAACGTGT	GTTAATGTGG	TATGTGGCAG	AGCTGATGGA	AAAGAGAGTG	13790
50	TG	TGTGTGAAT	TTTTTTTTCT	GAGAAAACCT	ACTGGAAGCA	AAATAAGTTG	GTCTTTACAG	CATATACAGC	13860
	AG	CAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAACACACA	CCAGCAACAG	AAATAAAACA	AAAGCAGTCA	13930
	AG	GAAGGGGA	GGTGAACGTT	CCCTGGTTTG	GTGTTGGGGA	AGGACACACA	GGGAGGCCGA	TGAACACAGT	14000
	GAG	GCAACCGG	GCATTGCTTT	CACTGCAGAG	AAACTCAGCT	TGCCTGAGCC	ACAGTGAAAA	TGGCCATTCC	14070
	CT	GAGGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTG	ACATTTCATC	TCTCACTTTG	14140
55	TT	CTCTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GC	AAAGGGCA	TGCATGATTG	CAGCCTGGCG	TCCTGCTCCG	GGGCCCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AA	GTGAGACC	CATAGGCTCA	GGGTGAGCCG	GAGGCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAAGAAA	14350
	TG	GACGTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AG	ACCCATCC	CTCAAGAGAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TT	TTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTGAGGCAA	GCTGGAAAGA	CTTTTCTGGA	AAGCAGCTTG	14560
	TT	TGCTGGA	AGTCTCTACA	ATGTCCTGTG	TCTTCCCACT	AATTCACCTT	CTGAAGTGAC	CAGACATTAT	14630
	CAC	GGGTCTT	ATTTACCAT	TCCAGTGTTT	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCCC	14700
	AA	ATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTG	CTCTGCCATT	AAACATTTTT	CAAGAATTTT	14770
	TT	GAGGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAAAGAAC	14840
65	AC	ACCCAGG	AGCCTGCCGT	GAATGTGATG	TGTGTTTATC	TTTGAGCATG	GACATACATG	GGCAGTGAGT	14910
	GG	TGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCCAT	CCTGCCCTCT	TGGAGACACC	ATGTGTGCCA	14980
	CG	TGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTCCATC	CCTGAGATTG	AAACACAGTG	15050
	AG	ATTTCCCC	CGCCCACTC	AGTGTCTCTC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCGAGGGAC	15120
	CG	CCGGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
70	GAT	GATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTTGGAATA	CACCTCAACT	CACAGGCCAG	GTCTGTGGTG	15260
	AG	TTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTTGG	AGTCCATGGA	GTGAGCACCC	AGCCCGCTCG	15330
	GG	CTGACAGC	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	CG	AGAGGGGG	GCTGGGTGTG	GGGCGAGGAC	CTGTGTCTGA	CATTCCTCCC	TGTGTCTCAG	CTATGCCCGG	15470
	AC	CTTCAATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
75	TT	GGGGTCTT	GCGGCTGAAG	TGTCACAGCC	TGTTTCTGGA	TTTGAGAGGT	AGCAGGCTGA	TGGTCAGCAC	15610
	AG	AGTTGCTA	GTTACAGGAG	TGTGTGCGCA	AGTATGTGTG	TGTGTGTGTG	CGCGCGTGCC	TGCAAGGCTG	15680
	AT	GGTGACTG	GCTGCACGTA	AGAGTGCACA	TGTACGCATA	TACACGTGAG	CACATACATG	TGTGCATGTG	15750
	TG	TACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCAACAAGT	TGTGCACATG	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCCCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCCAGGCC	TGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCTGTGG	GCTTCTGTGT	CCACTCCCC	16100
5	TCTCTGTGG	GCATTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTGT	16170
	GTGGGCATCT	GGTCCACCT	CCCCTCTCTG	TGGGCATTTC	CGTCCACTCC	CTCTCTGTGT	TCCTCTGTGT	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTGCCG	16310
	GGTGAGGGCC	AGGCCGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCCCT	TTTCTGTGTC	16380
10	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGGCG	ACTCTAGGTG	AACAGCCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGTGTGA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	GAGGCCAGC	CTCCAGGGAC	CCTCCGCGCT	CTGCTCACCT	16590
	CTGACCCGGG	GCTTACCTTT	GGAACTCCTG	GGTTTTAGGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCCCTGT	GCACAGTTCT	GTTCCGCTGG	CTCTGTGCAA	ASCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
	AGGAGCCGGT	GTGGCCCCAG	GTGTCCCCAC	TGTGCTGTG	CACTGGCCGT	GGGACGTCT	GGAGGCCATC	16800
15	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGTGT	16870
	CTGAGGAGAT	GATGGGAGAT	CAGATGCCCG	GAGGATTTGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCCACACA	GCCCGGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCTT	GGAACTTCC	CTGTCTTGG	TGGTACGGGG	GTGCCCTCTG	17080
20	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCCTGG	17150
	AGTCAGGGCA	GGTGGTGGCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGCACACCA	GGCCCGGGCC	17220
	TCCACCTCAA	AGGGCCTCCC	GAGCCACTGG	CAGGTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATAGAG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTATC	CCAGGGCCGA	GGCTGCGCGA	ATTACCGTGC	ACACTTGATG	17360
	TGAATGAGG	TGCTGCTCTA	TCGTGGAAGG	CCAGCAAGGG	CTACGGGAG	AGTTTTCCAT	TACAAGGTGC	17430
25	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTGCGCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
	TGCATGTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCCAG	GGTGCCTCCG	17570
	GCTCAGACCG	CCCTCCTCTC	TGCTTCTCTT	CTCTGCCTCA	AATCTTCCCT	CGTTTGATC	TCCCTGACGC	17640
	GTGCTGGGGC	CCTCGTGCAA	GCTGCTTGAC	TCCTTTCCGG	AAACCTTGG	GGTGTGCTGG	ATACAGGTGC	17710
	CACTGAGGAC	TGGAGGTGTC	TGACACTGTG	GTGACCCCA	GGGTCCAGCT	GGCGTGCTTG	GGGCCTCCTT	17780
30	GGGCCATGAT	GAGGTCAGAG	GAGTTTTCCC	AGGTGAAAC	TCCTGGGAAA	CTCCAGGGC	CATGTGACCT	17850
	GCCACCTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGTCTCTA	TTTCCCCACC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	GTCGCTTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCGAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
35	TTTACAGGCA	GCCAGGCTGC	AGTGGGCGAG	CTGTGGTGG	TCCACGTGGC	GCTGGGGGGG	GGGTCTGATT	18200
	CAAAATCGCT	GGGGCTCGGC	CTTCTTGGCC	CGTGTGGGCC	GGCCCTCCAC	ACGGGCTTGG	GGTGGAGGCC	18270
	CCGACCTCTA	GCAAGTGGCT	ATTTCTCCCT	TTGGAAAGAGA	GGCCCTCACC	CATGCTAGGT	GTTTCCCTCC	18340
	TGGGTACAGG	CGGTGGCCGT	GTGGCAACCC	CGGGACCTTA	GGCTTATTTA	TTTGTTTAAA	AACATTTCTG	18410
	GCCTGGCCTT	CGTTTGTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGTGCTG	CTTTGACTGG	TGTGATCTCA	GCTCATTTCA	GAAGTGGCTC	AGGAAGTCAG	TGAGACCAGG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGC	TCAGGCACCT	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAGGTACA	18690
	CGGGGGGTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGTCTGTG	18760
	CACACCTGCT	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGTGGGCT	18830
45	CACACCTGTA	GTCCCAGCAC	TTTGGGAGGC	CGAGGGCAGA	GGATCCCTTG	AGCCCCAGGAG	TTTAAAGCCA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAATAA	AAACAAATAA	TTAGCTGAAC	ATGGTGGTGT	18970
	CGCCCTCTAG	TTCCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGA	CCAGGAGGT	GGAAGCTGCA	19040
	GTGAGCTGAG	ATTGCACCAC	TGTACTGCAG	CCTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACAACAAA	19110
	GAAGACTGAC	AAATGCAGTT	TCTTGGAAAG	AAACATTTAG	TAGGAACCTA	ACCTACACAC	AGAAGCCAG	19180
50	TCGGTGTCTC	GGTGTCAAGT	AGATGAGATG	ATGGGTCCCT	ACACCATCAC	CCCAGACCCA	GGGTTTATGC	19250
	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTTGTCTGAC	19320
	GAAGGGGAGC	ATTCAATGATA	AGTACTCTGT	GGTACACAAG	GAACAATGGA	TAAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAGGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTACGCC	19460
	TCCACATGCG	TGTTTATACA	GATGGTGCAC	AGAAACGCAG	TGTACCTGTG	CACACACAGA	CACGCAGCTA	19530
55	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGCAATGA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGATTCTA	TGCACGCACA	CAGGCACCCG	TGGGCCCATG	CCCACACCCA	CGAGCACCGT	19670
	CTGATTAGGA	GGCCTTTCTT	CTGACGCTGT	CCGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTCATC	AGCAAGTTTG	GAAGAACCCC	ACATTTTCTC	TGCGGCTCAT	CTCTGACACG	GCCTCCCTCT	19810
	GCTACTCCAT	CCTGAAAGCC	AAGAACGCAG	GTATGTGCAG	GTGCTGGGCC	TCAGTGGCAG	CAGTGCCTGC	19880
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAAT	TCTTACCCCT	TTTCGATCA	19950
	GGAAAGTGGT	TAACCCAAAC	ACTGTCAAGC	TCGTCTGCCC	GGCCTCTCGT	GGGGTGGGCA	GAGCACCTGA	20020
	TGGAAGGGAG	AGGAGCTGTC	TGGGAGCTGC	CATCCTTCCC	ACCTTGCTCT	GCCTGGGGAA	GGCGTGGGGG	20090
	GCCTGGTCTC	TCTGTGTTGC	CCCATGGTGG	GATTTGGGGG	GCCTGGCCTC	TCTGTGTTGC	CCTGTGGTGG	20160
	GATTGGGCTG	TCTCCCGTCC	ATGGCATTAA	GGGCCCTTGT	GCAAAACCCAG	GCCAAGGGCT	TAGGAGGAGG	20230
65	CCAGGCCAGG	GCTACCCAC	CCCTCTCAGG	AGCAGAGGCC	CGGTATACCC	ACGACAGAGC	CCCGCGCGGT	20300
	CCTCTGCTTC	CCAGTCACCG	TCTCTGCCCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTTA	AGCCATGTG	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTCTGTGTC	TTTCTGTGTT	20440
	GTGGAATTTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTGGGAGC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCG	GTGTCTCTTG	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCGCCCA	GGGATGTGCG	TGGGGGCCAA	GGGCGCCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCAGTGG	CTGTGCCACC	AAGCATTCTT	GCTCAAGCTG	ACTCGACACC	20720
	GGTGTCACTA	CGTGCCACTC	CTGGGGTCA	TCAGGACAGG	CAAGTGTGGG	TGGAGGCCAG	TGCGGGCCCC	20790
	ACCTGCCGAG	GGGTATCCTT	TGAACGCCCT	GTGTGGGGCG	AGCAGCCTCA	GATGCTGCTG	AAGTGCAGAC	20860
	GCCCCCGGGC	CTGACCTTGG	GGGCCCTGGG	CCACGCTGGC	AGCCCTATGT	GATTAAACGC	TGGTGTCCCC	20930
75	AGGCCACGGA	GCCTGGCAGG	GTCCCCAATC	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGGC	GATGGCTCCC	21000
	CACGCTTGGG	AGCCTTCTGA	CCCTGACCTT	GTGTCTCTCT	ACAGCCTCTT	CCCTGGGTGC	TGCCCTGAGC	21070
	TCTGGGGTGC	CTGAGCAAGT	TCTCTCCCGC	CCCGCCGCTT	CCAGCGTCAC	TGGGCTGCCT	GTCTGCTCGC	21140
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTCCCAAC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21210

	TGCCCCGCCA	CCCACACGTC	CTAGGAGGGT	TGGAGGATGC	CACCTCTGGC	CTCTTCTGGA	ACGGAGTCTG	21280
	ATTTTGGCCC	CGCAGCCCAG	ACGCAGCTGA	GTCGGAAGCT	CCCCGGGACG	ACGCTGACTG	CCCTGGAGGC	21350
	CGCAGCCAAAC	CCGGCACTGC	CCTCAGACTT	CAAGACCATC	CTGGACTGAT	GGCCACCCCG	CCACAGCCAG	21420
	GCCGAGAGCA	GACACCAGCA	GCCCTGTCAC	GCCGGGCTCT	ACGTCCCAAG	GAGGGAGGGG	CGGCCACAC	21490
5	CCAGGCCCCG	ACCGCTGGGA	GTCTGAGGCC	TGAGTGAGTG	TTTGGCCGAG	GCCTGCATGT	CCGGCTGAAG	21560
	GCTGAGTGTC	CGGCTGAGGC	CTGAGCGAGT	GTCCAGCCAA	GGGCTGAGTG	TCCAGCACAC	CTGCCGTCTT	21630
	CACTTCCCCA	CAGGCTGGCG	CTCGGCTCCA	CCCCAGGGCC	AGCTTTTCTT	CACCAGGAGC	CCGGCTTCCA	21700
	CTCCCCACAT	AGGAATAGTC	CATCCCCAGA	TTCGCCATTG	TTCAACCCCTC	GCCCTGCCCT	CCTTTGCCCT	21770
	CCACCCCCAC	CATCCAGGTG	GAGACCTGA	GAAGGACCTT	GGGAGCTCTG	GGAATTTGGA	GTGACCAAG	21840
10	CTGTGCCCTG	TACACAGGCG	AGGACCTGCG	ACCTGGATGG	GGGTCCCTGT	GGGTCAAAT	GGGGGGAGGT	21910
	GCTGTGGGAG	TAAAACTAGT	AATATATGAG	TTTTTCAGTT	TTGAAAAAAA	TCTCATGTTT	GAATCTTAAT	21980
	GTGCACTGCA	TAGACACCAC	TGTATGCAAT	TACAGAAGCC	TGTGAGTGAA	CGGGGTGGTG	GTCACTGCCG	22050
	GGCCATGGCC	TGGCTGTGCA	TTTACGGAAG	TCTATGAGTG	AATGGGGTTG	TGGTCAGTGC	GGGCCCCATG	22120
	CCTGGCTGGG	CCTGGGAGGT	TTCTGATGCT	GTGAGGCAGG	AGGGGAAGGA	GGGTAGGGGA	TAGACAGTGG	22190
15	GAGCCCCCAC	CCTGGAAGAC	ATAACAGTAA	GTCCAGGGCC	GAAGGGCAGC	AGGGATGCTG	GGGGCCACGC	22260
	TTGGGCGGGG	GGGATGATGG	AGGGCCTGGC	CAGGGTGGCA	GGGATGATGG	GGGCCCCAGC	TGGGGTGGCA	22330
	GGGGTGATGG	GGGGGGCTGG	TCTGGGTGGC	GGGGAAGATG	GGGAAGCCTG	GCTGGGCCCC	CTCCTCCCCC	22400
	GGCTCCACAC	TGCAGCCGTG	GATCCCGATG	TGCTTCCCTG	GTGCACATCC	TCTGGGCCAT	CAGCTTTTCA	22470
	GGAGGTGGGG	GGCAGGGGCA	TGACACCATC	CTGTATAAAA	TCCAGGATTC	CTCCTCTCTG	ACGCCCCAAAC	22540
20	TCAGGTTGAA	AGTCACATTC	CGCCTCTGGC	CATTCTCTTA	AGAGTAGACC	AGGATTCTGA	TCTCTGAAGG	22610
	TGGGGTAGGG	TGGGGCAGTG	GAGGGTGTGG	ACACAGGAGG	CTTCAGGGTG	GGGCTGGTGA	TGCTCTCTCA	22680
	TCTCTTATCT	ATCTCCCACT	CTCATCTCTC	ATCCTCTTAT	CATCTCCCACT	TCTCATCTGT	CTCTCTCTTA	22750
	TCTCCCACTC	TCTCTCTTAC	ATCTCCCACT	CTCATCTCTT	ATCCTCTTAT	CTCCTAGTCT	CTCCTAGTCT	22820
	CATCCAGACT	TACCTCCCACT	GGCGGGTGCC	AGGCTCCGAG	TGGAGCTGGA	CATACGTCCT	TCTCAGGCA	22890
25	GAAAGAACTG	GAAGGATTGC	AGAGAACAGG	AGGGGGGGCT	CAGAGGGAGC	CAGTCTTGGG	GTGAAGAAAC	22960
	AGCCCTCTCT	CAGAAGTTGG	CTTGGGCCAC	ACGAAACCGA	GGGCCCTGCG	TGAGTGGCTC	CAGAGCTTTC	23030
	CAGCAGGTCC	CTGGTGGGGC	CTTATGGTAT	GGCCGGGTCC	TACTGAGTGC	ACCTTGGACA	GGGCTTCTGG	23100
	TTTGAGTGCA	GCCCGGACGT	GCCTGGTGTC	GGGGTGGGGG	CTTATGGCCA	CTGGATATGG	CGTCATTTAT	23170
	TGCTGTCTGT	TCAGAGAATG	TCTGAGTGAC	CGAGCCTAAT	GTGTATGGTG	GGCCCAAGTC	CACAGACTGT	23240
30	GTCTGAAATG	CACCTCTGGT	CCTGGAGCCC	CCGTATAGGA	GCTGTGAGGA	AGGAGGGGGT	CTTGGCAGCC	23310
	GGCCTGGGGG	CGCCTTTGCC	CTGCAAACTG	GAAGGGAGCG	GGCCCGGGCG	CCGTGGGGCG	ACGACCTCAA	23380
	GTGAGAGGTT	GGACAGAAAC	GGCGGGGGAC	TTCACAGGAG	CAGAGGGCCG	TGCTCAGGCA	CACCTGGGTT	23450
	TGAATCAGAG	ACCAACAGGT	CAGGCCATTG	TTCAGCTATC	CATCTTCTAC	AAAGCTCCAG	ATTCCTGTTT	23520
	CTCCGGGTGT	TTTTTGTGTA	AATTTTACTC	AGGATTACTT	ATATTTTGTG	CTAAAGTATT	AGACCTTTAA	23590
35	AAAAGGTATT	TGCTTTGATA	TGGCTTAACT	CACATAAGCAC	CTACTTTATT	TGCTGTGTTT	TATTTATTAT	23660
	TATTATTATT	ATTAGAGATG	GTGTCTACTC	TGTCAACCCAG	GTGTGTTAGTG	CAGTGGGCACA	GTCTGGGCTC	23730
	GGTGTAGCCG	CAAAACCCCA	GGCTCAAGTG	ATCCTCCGGC	CTCAGCTTCC	CAGAGTGCTG	GGATTACAGG	23800
	TGTGACCCAC	TGCCCTTGCC	TGGCACTTTT	AAAAACCACT	ATGTAAGGTC	AGGTCCAGTG	GCTTCCACAC	23870
	CTGTCACTCC	AGTAGTTTGG	GAAGCCGAGG	CAGAAGGATT	GTCTGAGGCC	AGGAGTTTGA	GACCAGCATG	23940
40	GGTAACATAG	GGAGACCCCA	TCTCTACAAA	AAATGCAAAA	AGTTATCCGG	GCGTGGGGTC	CAGCATCTGT	24010
	AGTCCCAGCT	GCTCGGGAGG	CTGAGTGGGA	GGATCGCTTG	AGCCCGGGAG	GTCATGGCTG	CAGTGAGCTG	24080
	TGATTGTACC	ATCGCACTCC	AGCCTGGGCA	ACAGAGTGAG	ACCCTGTCTC	AAAAAAAAAA	AAAAAAAAAG	24150
	AAGGAGAAGG	AGAAGAGAAG	AAGAAGGAAG	AAGGAAGAG	AAGAAGGAAG	AAGAAGGAAG	AAGAAGGAAG	24220
	AAGGAGGCCCT	GCTAGGTGCT	AGGTAGACTG	TCAAACTCA	GAGCAAAATG	AAAATAACAA	AGTTTTAAAG	24290
45	GGAAAGAAAA	ACCCACAGTC	TTTGGACTTC	CTTAGGCTCG	AACTTCACTC	CAAGCAGCTT	CCTTCCACAG	24360
	ACAAGCGTGT	ATGGAGCGAG	TGAGTTCAAA	GCAGAAAGGG	AGGAGAAGCA	GGCAAGGGTG	GAGGCTGTGG	24430
	GTGACACCA	CCAGGACCCC	TGAAAGGGAG	TGGTTGTTTT	CCTGCCTCAG	CCCCACGCTC	CTGCCGGTCC	24500
	TGCACCTGCT	GTAACCGTCG	ATGTTGGTGC	CAGGTGCCCA	CCTGGGAAGG	ATGCTGTGCA	GGGGGCTTGC	24570
	CAAACTTTGG	TGGGTTTCAG	AAGCCCCAGG	CACCTTGTGG	AGGCACAATT	ACAGCCCCCTC	CCCAAGATG	24640
50	CCCACGTCTC	TCTCCTGGAA	CCTGTGAATG	TGTCAACCCG	AAGGCAGAGG	CTGGTGAAGG	CTGCAGGTGG	24710
	AATCAGGGCT	GCCAGTCAGC	CGATCTTAAG	GTCATCCTGG	ATTATCTGGT	GGGCCTGATA	TGGCCACAG	24780
	GGTCCCTAGA	AGTGAGAGAG	GGAGGCAGGG	GAGAGTCAGA	GAGGGGACGT	GAGAAGGACC	ACTGGCCACT	24850
	GCTGGCTTTG	AGATGGAGGA	GGGGGTCCCC	AGCCAAAGGA	TGGGGGCAGC	CGCTCCATGC	TGGAAAAGCA	24920
55	AGCAATCTCT	CCCGGTCTCT	AGGGCACACG	GGCCTGCCCA	CGCCTCGATT	TCAGGCCAGT	GGGACCTGTT	24990
	TCAGCTTTCC	GGCCTCCAGA	GCTGTAAGAT	GATGCGTTTG	TGTTACGCCA	CTAAGCTGCA	GTGATTCTGT	25060
	ACAGCAGCAA	ATGGAATAGC	AGTACAGGGA	AATGAATACA	GGGACAGTTC	TCAGAGTGAC	TCTCAGCCCA	25130
	CCCTGGG							25138

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon			
Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A G G
Häufigk.(%)	70	60	80		100	100	95	70		80	100	100 60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCTCCCCGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCTCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTGATCTCTGCC
TCTGCTCTCCCTCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTCACGTTTGTATGGACACGCGGTTTCCAGGCGC
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGTTGCCGGCAATGGGGAGAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCTCTCGCAATTCAAGGGTGGGAATGAGAGGTGGGA
CGAGAACCCCTCTTCTGGGGTGGGAGGTAAGGGTTTTCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCAGCACTTTGGGAAGCTGAGGCAGGTGGA
TCACCTGAGGTGAGGATTGAGACCAGCTGACCAACATGGTGAAACCTATCTGTACTAAAAATACAAAATTAGCTG
15 GGCATGGTGGTGTGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAACTCTGTCTTTAAAAAAAAGTGTT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCATCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGACAGTGTTTGTGGGTGTTAGGGG
ATGGTGTCTGCTGGGCCCTGCCGTGTCCCAACCTGTTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCCTT
20 TGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCTCCCACTCCCAAGAC
ATGTAAGACTTCGGCCATGCAGACAAGGAGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTCTTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGTAACCTCGGCGGTGTTTA
CAGCAGGTGCTTGAATGCTGCGTCTTGGCTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCCGAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTGTGCTCCAGTTCCTTCGTTGAG
25 GAGAGTTGAGTTCTCTGATCAGGACTCTGCCTGTCTGCTGTTCTCTGACTTCAGATGAGGTACAACTGCCCCCTGG
CTTATGACAGGGAGTGAGGCGTGGTCCCCGGGTGTCCTGTGCAGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTACAGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCGGGTGT
CCCTGTGCAGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTGCAGTGTAGGGTGAGTGAGGCGTGGTCCCCGG
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTGCAGTGCAGGGTGAGTGAGGCGCGGTCC
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTGCAGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTGCAGTGCAGGGTGAGTGAGG
CGCGGCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCAAGGTGTCCTTGGCGTTTGTCTCACTTGAGCTTGCTCCTGAATGTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTGTGCCATTGCTGGGTAGATGGTGCAGGCGCAGTGTGGTCCCAAGCCTATCTTTCTGATGCTCGGCTCT
35 TCTTGGTCACTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGCGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTTGTCTACCACGTGCCCGCCACATGCATGCTGCCAATACTCC
TCTCCAGCTTGTCTCATGCCGAGGCTGGACTCTGGGCTGCCTGTGTCTGCTGCCAGTGTGCTGGAGACATCCAGAA
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTACCCAGCCCCCTCACTGTCTGTTTTCTCCCAAGTGCCCTCTGTC
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCT
40 CCAGTGCAGCCCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGTCTCTGTCTGCTCTGCTGCTGAGACCCACGTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTCAGACTTCCCTCTTGGGTCTTAGTTTGAATTTCACTGATTTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTTCTTGGTTTATTCTTCATTCCCTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTCAAGTGTCTTAAA
ATACTTCAAAGTGTTAATACTTCTTTAAGTATTCTTATTCTGTGATTTTTTTCTTGTGCACGCTGTGTTTTGACGTGA
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATAACCAATTATTTGAAGTTTGGCGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTCTAGATGCATGAAATTTCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTCATT
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATCTTTTTTAAATGTCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTTCTGTGTGTCTGTTTTCTGCCTTTAATTTATATATATATATATATTTTTTTTTTTTTTGGAGACA
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGAGTGGTGTGATCACAGGTCAGTGTAACTTTTACCTTCTGCGCTGAGCCGT
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAAGTCAGACACGACCGCTACACCTGGCTAATTTTAAATTTTTTCTGGA
15 GACAGGGTCTGTGTGTTGCCAGGCTGGTCTCAAACCTCTGGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTCTGTTA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCGTTGTTTAACTGGATAACCTGATTTATTTTCATTTTTTGTCT
ACTAGAGACCCGCTGGTGCACTCTGATTCTCCACTTGCCTGTGTGATGTCTCGTTCCCTTGTCTCTCACCACCTCTTG
GGTGGCATGTGCGTTTCTGCGGAGTGTGTGTGATCCTCTCGTGTGCTCCTGGTCACTGGGCATTGTCTTTATTTCT
20 CTTTGCTTAGTGTTACCCCTGATCTTTTATTGTCTGTTGTTGCTTTTGTATTAGACAGTCTCACTCTGTCAACCA
GGCTGGAGTGTAAATGGCACAATCTCGGCTCACTGCAACCTCTGCCTCCTCGGTTCAAGCAGTTCTCATTCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCACCAACGCTGGCTAATTTTGTATTTTGTAGTAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGATCTGCCCGCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCTTGTCTGAGCAATAAGACCTT
25 AGTGATTTTGTAGCTCTGGCCACCCCGAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTTCTATCTCAGGCATCTTGACA
CCCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTGAGTGTCTGTAGCTTTGCCCGCCCTGCTTTTCTCTC
TTTGTCTCCCGTCTGTCTTCTGTCTCAGGCCCGCGTCTGGGGTCCCTTCTTGTCTTTGCGTGGTTCTTCTGTCTTG
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCTCCATGGCATCTAGCGACGTCGGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCAGGAGGAGGGCGGTATCTTGGCCCGTGTGCTGTGGAGCACCAGTGGCCAGCGTTC
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCTGGGTTTCAAGCTGGAAAACCCAGGCATGTGGGGTCTGGTGGCT
CCGCGGTGTGAGTTTGAATCGCGCAAACCTCGGCTGTGGCGCCAGCTCTGACGGTGTGCTGCGGGGGAGTGTCTG
CTTCTCTCTTCTGCTTGGGAACCAAGGATGAGGCTCCGAGCCGTTGTGCGCCAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCTGTAATCCAGCACTTTGGGAGGCCAAGGCGGG
TGGATCACGAGGTGAGAGGTGAGACCATCTGGCCAACATGATGAAACCCATCTGTACTAAAAACAAAAATTAGC
35 TGGGCGTGGTGGCGGGTGCCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGCAGTGAGCCGACATTGCACCACTGCACTCCAGCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA
AAAAAAAAAAAAATTCTAGTAGCCACATTAAAAAAGTAAAAAAGAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATCATTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTGTAGC
TTTGTCTGCGGGATCCCGTGTGTAGGTCCCGTGGCTGGCCATCTCGGCTGGACCTGCTGGGCTTCCCATGGCCATGGCT
40 GTTGTACCAGATGGTGCAGGTCCGGGATGAGGTCCGACGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGGATGGTGCAGG
TCTGGGATGAGGTCCGACGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTCTCCAG

5 G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C T G C T G T G A G C T G G A T G
T G T G G T G T C T G G A T G G T G C A G G T C A G G G T G A G G T C T C A G G C C C T C G G T A A G C T G G A G G T A T G G A G T C C G G A T G A T G C A
G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G G T G A G G T C A C C
A G G C C C T G C G G T G A G C T G G G T G T C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G A C G G T G C C A G A C C A T G C
10 G G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C T G G G T G A G G T T G C C A G G C C C T G C T G T G A G T T G G A T G T G G G G T
G T C C G G A T G T G C A G G T C C G G T G T G A G G T C A C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T
G G G G T G A A G G T G C C A G G C C C T G C T T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T G C C C A G
G C C C T C G G T G A G C T G G A T G T G C A G T G T C C A G A T G G T G C A G G T C C G G G T G A G G T C G C C A G A C C C T G C G G T G A G C T G G A T G
T G C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G G C C C T C G G T G A G C T G G A T G T A T G G A G T C C G G A T G G T G C C
15 G G T C C G G G T G A G G T C G C C A G A C C C T G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T A C A G G T C T G G A G T G A G G T G C C C
A G A C C C T G C T G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C A G G G T G A G G T C T C A G G C C C T C G G T G A G C T G G A
G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A A C T G G A T G T G C G G C G T C T G G A T G G T
G C A G G T C T G G G T G T G G T G C C A G G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G
C C A G G C C C T G C T G T G A G C T G G A T G T G C G G C G T C T G G A T G G T G C A G G T C T G G G T G T G G T G C C A G G C C C T C G G T G A G C T G
20 G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T T G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G
G T G C A G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G G T G A G G T
C A C C A G G C C C T G C G G T G A G C T G G T T G T G C G G T G T C C G G T T G C T G C A G G T C C G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T G T G C G G T G T C C C C G T G T C C G G A T G G T G C A G G T C C A G G G T G A G G T C G C T A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G T G A G G T C G C C A G G C C C T T G G T G A G C T G G A T G T G C G G T G T C T G C A T G G T G C A G G T C T G
25 G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G T G G T G T C C G G A T G G T G C A G G T C C G G C G T G A G G T C G C C A G G C C C T
G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G T G A G G T A G C C A A G G C C T T C G G T G A G C T G G A T G T G G G
G T G T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T
C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G
C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C A G T G A G C T G G A T G
30 T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C G G A T G G T G C A
G G T C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C
A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G
A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G A T G T G C A G T G T A C G G A T G G
T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T G G G C T G T A T G T G T T G T C T G G A T G G T G C A G G T C C G G G T G A G G T
35 C G C C A G G C C C T G C G G T G A G C T G G A T G T G T G G T G T C T G G A T G C T G C A G G T C C G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T A T G C G G T G T C C C C G T G T C C G A A T G G T G C A G G T C C A G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G T G A G G T C G C C A G G C C C T T G G T G A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C C G
G G G T G A G G T C A C C A G G C C C T C G G T G A T C T G G A T G T G G C A T G T C C T T C T C G T T A A G

35 Intron 3 (SEQ ID NO 6)

40 G T A C T G T A T C C C C A G C C A G G C C T C T G C T T C T C G A A G T C C T G G A A C A C C A G C C C G G C C T C A G C A T G C G C C T G T C T C C A C T
T G C C T G T G C T T C C C T G G C T G T G C A G C T C T G G G C T G G G A G C C A G G G G C C C G T C A C A G G C C T G S T C C A A G T G G A T T C T G T G
C A A G G C T C T G A C T G C C T G G A G C T C A G G T T C T C T T A C T T G T A A A A T C A G A G T T T G T G C C A A G T G G T C T C T A G G G T T T G T A
A A G C A G A A G G G A T T T A A A T T A G A T G G A A C A C T A C C A C T A G C C T C C T T G C C T T T C C C T G G G A T G T G G G T C T G A T T C T C T C
T C T C T T T T T T T T C T T T T T G A G A T G G A G T C T C A C T C T G T T G C C C A G G C T G G A G T G C A G T G G C A T A A T C T T G G C T C A C T

GCAACCTCCACCTCCTGGGTTTAAGCGATTACCCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
 GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
 TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
 CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTAGGTATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG
 5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCACCATGAGCTAACTTCTAGGTGGCTGCATTGTAATG
 GCTGTGAGATTTTGTCTGCAATGTTGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGTCATCAGTGAG
 GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCAGGCCATGGTATTAGCTTCTCCGTGTCCGCCCAGGC
 TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGGCTTGTGAC
 ACCCATGCCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
 10 GACCCGACGTGGTGTGGGGCCATTTCCTTGTCATCTGGGGAGGGTCAGGGCTTCCCTGTGGGAACAAGTTAATACAC
 AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGGAAAGAAT
 TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCCAAGATGCTCCTGTGCACTACTGGGACTGTTGTCTG
 CCTGGGGGGCTTGGAGGCCCTCCTCCCTGGACAGGGTACCGTGCTTTTCTACTCTGTGGCCTGCGGCTGCGGTC
 AGGGCACCAGCTCCGGAGCACCCGCGGCCCCAGTGTCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT
 15 GTGCGCGCTCCAGCCCCCTGCCCATGGTGGTTTTGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGCTGTGGG
 CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCGCTCTCCATCTGAAGGATGTGGCT
 CTTTCTACCTGGGGGTCTGCTGGGGCCAGCCTTGGGCTACCCAGTGGCTGTACCAGAGGGACAGGCATCTGTGTGG
 AGGGGCATGGGTTACAGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTCAACCTGGGG
 GTTGACCGCCGGACTGGGCGTCCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
 20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
 TGAGTCGGTGGGGGCTGTGGCTTCCCGTGAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
 CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTTTAAACAGAAAGTGGCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
 CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCAGCGCCAAACCATTTGTGCGCACAGTAGGTTGGCCGAGG
 TGCCGGTGCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
 CAGCCGGGCCAGGGCTGGATGCAGCAGGCCCGAGGTCTGGATCCGTGTCTGCTGTGGTGCGCAGCCTCCGTGCGCT
 TCCGCTTACGGGGCCCGGGACAGGCCACGACTGCCAGGAGCCACCGGGCTCTGAGGATCCTGGACCTTGCCCCACGG
 30 CTCCTGCACCCCCACCTGTGGCTGCGGTGGCTGCGGTACCCCGTCATCTGAGGAGAGTGTGGGGTGAGGTGGACAGAG
 GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
 TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGGTGGGCAC
 TTGGCCGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCCGCCAG

35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
 GAGGTACTCTGGGTGGGCCGAGGGAGTGACGTTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
 CTTACGCTTTCTGTCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGCTCCTATTTCCCAAGGAGGGTCCCACTG
 GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCGGTGCCTTGACCCCACTCTGAGCCAG
 40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCGCCTGCCCTTGGGGTCTGGAGTGGTGGGGTGCAGAGAGAG

5 GTAAGGTTACAGTGTGATAGTCGTGTCCAGGATGTGTGTCTCTGGGATATGAATGTGTCTAGAAATGCAGTCGTGTCTGTG
ATGCGTTTCTGTGGTGGAGGTACTTCCATGATTTACACATCTGTGATATGCGTGTGTGGCACGTGTGTGTGTGCGTGGTGCAT
GTATCTGTGGCGTGCATATTGTGGTGTGTGTGTGTGTGGCACGTGTGTGTCCATGGTGTGTGTGCCTGTGGTGTGCATG
TGTGTGTGTCTGTGACACGTGCATGTTTCATGCTGTGTGCTGCATGTCTGTGATGTGCCTATTGTGGTGTGTGTGTGCAT
GTGTCCGTGACATATGCGTGTCTATGGCATGGGTGTGTGTGGCCCCCTTGGCCTTACTCCTTCCTCCTCCAGGCATGGTCC
10 GCACCATTTGCTCTCACGCTCTCGGGTGTGGTTTGGGGAGCTCCACATTACGGGTCTCCTTCTAGCATGGGTGCCCTT
GTCCTGTCAAGGGCTGGGCCTTGGAGACTGTAAGCCAGGTTTGAGAGGAGAGTAGGGATGCTGGTGTACCTTCCTGGA
CCCCCTGGCACCCCCAGGACCCAGTCTGGCCTATGCCGGCTCCATGAGATATAGGAAGGCTGATTAGGCCTCGCTCCCC
GGGACACACTCCTCCAGAGCGGCCGGGGCCCTTGGGGCTCGGCAGGGGTGAAGGGGCCCTGGGCTTGGGTTCCACCC
AGTGGTCATGAGCACGCTGGAGGGGTAAACCTCAAAGTCGTGCCAGGCCGGGGTGACAGAGGTGAAGAAGTATCCCTGGA
15 GCTTCGGTCTGGGGAGAGGCACATGTGGAAACCCACAAGGACCTCTTTCTCTGACTTCTTGAGCT

GTGTGGGATTGGTTTTTCATGTGTGGGATAGGTGGGGATCTGTGGGATTGGTTTTTATGAGTGGGGTAACACAGAGTTCAAG
 GCGAGCTTTCTTCTGTAGTGGGTCTGCAGGTGCTCCAACAGCTTTATTGAGGAGACCATATCTTCTTTGAACTATGGT
 CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCCTCCCCTGGGCTCCCTGTTCTGTTTCTTCCACTCTGGGGTCGTGTGGT
 CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCCTCCTTGTGTTTCATTGGCCTGGATGTGGCCCTGGCTACGCT
 CCGTCTTGGAAATCCCCCTGCGAGTTGGAGGCTTCTTCTTCTTTTTTCTTCTTTTTTTTTTTTGTATAACAGA
 GTCTCGCTCTTTTTTGCCAGGCTGGAGTGGTTTGGCGTGATCTTGGCTCACTGCAACCTGTGCTTCTGAGTTCAAGCA
 ATTCTCTTGCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCACCACCATGTGACTAATTTTGTAAATTTTAGTAG
 AGACGAGGTTTCTCCATGTTGGCCAGGCTGGTCTCGAACTCTGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAAAGT
 GCTGGGATGACAGGTGTGAACCGCGCGCCGGCCGAGACTCGCTTCTGCACTTCCGTGAGATCTGCAGCGATAGCTG
 CCTGCAGCCTTGGTGTGACAACCTCCGTTTTCTTCTCAGGTCTCGCTAGGGGTCTTTCCATTTCAAGTCTCTTCA
 CAGAAGAGTTTACGTGTGCTGATTTCCCGGCTGTTTCTGCGTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCEA
 TTTCTTTTAGGCTTTGTTTATTGTTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACTTTCTTT
 TCTAAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAAGCAGGGATCCCGAGGCCCTTGGCTGTGGAGTGGCACCGGTCT
 GGGGCTGTTAGGAACCCGGCGCACAGCGGAGGCTAGGTGGGTGTGGGGAGCCAGCGTTCCCGCTGAGCCCCGCCCC
 TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT
 GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTTGCTCCCAAAACCATCCCTTCCCCACTGCTG
 TCCTGTGGAATAATCGTCTTCCACGAACACAGTCCCTGGTACCACAATGGTTGGGGACCTGTGCTAAAGACCTGCTTCA
 GCAGCCTCTCGTCACTGTTGATATATTGGCTTTTCTGTGTTAGTCCAGAATAATTACGGATTCTGTGATGCTTCCGC
 CGACCTCAGACCCATGGGCTATTGTGGGCGTGTTCCTGCTCCTGGGTGGGAAGGGTGCAGGCCCCATGTACCTTCTCT
 GTTACTGCCTTCCAGGTTGGTCTCAGGGTTGAATCGTACTCGATGTGGTTTTAGCCACAGGCCCTGCCGCCAGCTCCTG
 GGGCTGGGGAACTGTGTAAGCACAGAGTACCGTGCAGCTCTTTGATGCCTCACAAGCTCGAGGCCTCCTGTGTCGG
 TGTTAGTGTGTGTACGTGCCTGCTCACATCTGTCTTGGGACGCAAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAGC
 GTCTGTGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCCGCTCTCTCTCCCGCTTCTCAGACTCTTCTCCTGCCTGTGCT

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTTGGTCACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGG
ATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTG
5 GGGTCTGATGTGGTGACTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATG
TGGTGACTGTGGATGGCAGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACT
GTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGG
CGGTCGTGGGGTCTGATGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTCA
CAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTACAG
10 GGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTTGGTCCCGGGG
TCTGATGTGTGGTGACTGTGGATGGCGATCGGTACAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCT
GATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGAT
GGCGGTTGGTCCCGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCAG
15 TCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGG
TCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGT
GGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGTGATCGGTACAGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGAT
GGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTAGGGTCTGATGTGTGGTGACTGTGGATGGCAGTCG
20 GTCACAGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGG
GGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGTGGTGACTGTGGATGGCGGTCGTGGGGTCTGAT
GTGGTGACTGTGGATGGTGATCGGTACAGGGGTCTGATGTGTGGTAGCTGCAGGTGGAGTCCAGGTGTGTCTGTAGCT
ACTTTGCGTCTCGCCCCCGCCCCCGTTTCCCAAACAGAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCCATCG
GGCTTGGCCGCAGGTCCACAGTCCCTGATCGGAAGAAACAAGTCCCAGCTCTGGCCGGGCAGGCCACATTTGTGGCTC
25 ATGCCCTCTCCTCTGCCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGGCAGTGGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTAGGCACCTGCTCACGTTGACTGCGCGGCCCTCTCTCCAGTT
CCGCACTGCCCTTTGTTTCATGATTGCTAAATGTCTTCTCTGCCAGTTTGTATCTTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTGAGCCGTGTCTGCCAGCTGGCCCTCAGTGCTGGGTCTGAGGCCAAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGCCGTGTTGAGCCACGCCCCGCTGAGCGGGCCTCTCAGTGCTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTGCAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTGCAGATGCCTGTAGCACTTGCTCGGC
35 TCTAGGGGACAGTCGTGTCCACGCATGAGGCTCAGAGACCTCTGGGCGAATTCCTTGGCTCCAGGGTGGGGGTGGAG
GTGGCCTGGGCTGTGGGACCCAGACCTGTGCCCGGCAGCTGGGCAGCAACTCCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCCATGTGTGTCTGCAGAGACTCGGCCCGCCAGCCACGATGGCCCTGCATTCCAGCCAGCCCCGCACTTCATCACA
AACACTGACCCCAAAGGGACGGAGGGTCTTGGCCACGTGGTCCTGCTGTCTCAGCACCCACCGGCTCACTCCCATGTG
40 TCTCCGCTCTGCTTTCGCAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCTGCCC
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC
5 TCCCGAGGCCCCGGAAACATGGCTCGGCTTGGCGCAGCCGGAGCGGAGCAGGTGCCACACAGGCCTGGAAATGGCAAGC
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCCGATACAAAA
GGATTTTATCCGATTCTCATTCTGTCCCTGTCGTGTGACCCCGGAGGGCGGGGCTCTTCTCTCTGTGACTAGATTT
CCCATCTGGAAAGTGCGGGGTGACCGTGTAGTTTGGCTCCTCTCGGGGGGCTGTGTGGCCATGGGGCAGGCGGCCTGG
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCTGGTGCCACATCACGTCCT
CTGGATTTTAAAGTAAACACACACCTCCCGGAGGCATCTGCCTGCGACCTGTGTGTGCTGGGAGAGTGGTAGCAC
GGAGGAAATTCGTGCACACTCAAGGTATCAGCAAGGTATCCGCAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTTATTTAAAAATATAACTATTAAATTATTGCATTATAAGT
AATCACTAATGGTATCAGCAATTATAATATTTATTAAGTATAATTAGAAATATTAAGTAGTACACAGTTCTGGAAAAA
15 CACAAATGACATGGCAGCAGAGTGAATTTTGGCCGAGGGACACGTGTGCACATGTGTGTAAGCGGCCCCCAGGCCCCAC
AGAATTCGCTGACAAAGTCACCTCCCCAGAGAAGCCACACGGGCCTCCTTCGTGGTTCGTGAATTTTATTAAGATGGATC
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGAAATGTGAGGTGATGACTGCGTCCTCATGCCCTGACAGACAGGA
GGTGACTGTGTCTGTCTCCTAGGACACGGACAGGCCCGAAGCTCTAGTCCCATCGTGGTCCAGTTTGGCCTCTGA
ATAAAACGTCTTCAAACCTGTGCCCCAAAACTAAGAACAGAGAGATTTCCCATCCCATGTGCTCACAGGGCGTA
20 TCTGCTTGGCTTGAATCGCTGGGCTGGCCGGACTCTAGAGTTGGTGGCTGTCTTCTGTGCAAAAAGTGAGTCTCTCT
GCCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTCTTTCTTTCTTTTCTTTTGGAGACGGAACGTCA
CTGTTGTCTGCTGGGCTTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGCCTCCCGGGTTCCAGCATTTCTC
CTGCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCCTGGCTAATTTTGTATTTTAGTAGAGAG
GGGTTTTTGCCATGTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGGAAGCCTCTTTTAAGGTGACCACCTATAGCGCTTCCCGAAAATAAC
AGGTCTGTGTTTTGAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGGATGGCTGAGGG
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTGTCTGAAAACGCA
CCCTTGGCATCCTGTTTGGAGAGTTTCTGCTTCTCGTTGGTCTATGCTGAACTAGGGGCAAGTTGTATCCGTTGGCGC
GCAGCGGCTACATGTAGGGTCAAGTCTTTTACCGTGGACAAATCCTTGAAAAAAAAAAGGAGTCCGGTTAAGCAT
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAATCTAAGATTAAAGAAACCTTAATGAAAGAAAACCTTGATGATTC
AGAGCAAGGATGTGGTACACCTGTGGCTGGATCTGTTTACGCCGCCAGTGCATGGTGAGAGTGGGAGCAGGGATTG
TTTGTTCAGAGGTCTCATCTGGTATGTTTCTGAGGTGTTTCCGGCTGAATGGTAGACGTGTCGTTGTGTGTATGAGGT
TCTGTGTCTGTGTGGCTCGGTTTGTAGGTACGCATGTCCAGCACATGCCCTGCCGCTCTCTCACCTGTGTCTTCCCGC
CCCAG

35

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTGATTGCTTTTGATGCATTCAAGTGTAAATATCTGGTGC
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACAGACAAGGTTGCAGCCCCTTCTTGGTATGAAGCCGACGGGAGGGG
TTGCACAGCCTGAGGACTGCGGGCTCCACGAGGCTCTGTCCAGCGGCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATCAACACCGAGGAAGCACACCAGCTTCTGTACGTACCCAGGTTT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTGCCCATCCCACCTTGCATGGGGTCTACACCCAAGGACGCACACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTAAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCTTACA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCCTG
TGTCCAAGTGTCTCATTGTTCAAGTTCACCTGTGAGTGAGAACATGTGGTGTGTTGGTTTTCTTTCCTTGCAATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACCTCATCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAAACATACGTGTCATGTGCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTCCACAATGGTTGAA
CTAGTTTACACTCCCAACAGTGTAAGTGTCTTGGTGTGGAGAGGATGTGGACAGCAGTTATTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTCTCTCGAAGAC
TCCGGGTTTTTCTGTGCATCTTTGAAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTCTCTTCTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAACAAAATTTCTTGACACACAACCTGTCTCTGGGATTTGGA
15 GGAAAGTGTCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCGCTGAGGGTCACACAGTGCACCATGCCAGCTTCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGGCGAGGTTCACAGCCCCAGCTTCTTACCGTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTCGTTCTGCTTCAGCTGGCACAGAATTGCACAA
GCTGTAGGTAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTTGAAA
20 GAAATTTAAGTTTTTCATTAAACCGCTTGGAGAATGTTACTTTATTTATGGCTGTGTAAATTGTTTGACATTCACTCCC
TCGTAGACAGATACTACGTAAAAAGTGTAAAGTTAACCTTGCTGTGATTTTCCCTATTTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCCGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG
25 CACCGCAGCGTTGTCTCTGCAAGTCTCTCTCTCTGCCGCTGTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCTTGGGGCGCAGGGGCACCTTCGGGAGGAGTGGGTACCGTGAGGCCCTGGTCTGTCAGAGACGCACCCAGGTT
ACACACGTGGTGAGTGCAGGCGGTGACCTGGCTCCTGCTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGCACAGGGCTGCAGGGCCGAGGCGGCAGCCTCCTCCCAGGCTGCACCTGAGCTGCGGA
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTCGCTGCGGTACGTTCTGCTGGGGTTGTTGGGATCGGTGGG
30 AGAATTTGGATTGCTGAGTGTCTGTCTTGAACCACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTGTGAAT
CAAATAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGGTGGGCTTGTTTAAAGTGCATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTCAGAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTGTGAAAACCATTTGGACCCGCCCTCCAAGTCCACCC
TCCAGGTCCACCCTCCAGGGCCGCCCTGGGCTGGGGTATGCCTGGCGTTCCTTGTGCCGAGCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGAACGCTGCTTCTGTGTGCAAGTTCCTGAGGGTGTGCGCAGGGAGGTGGCTCAGA
GTGTATGTTGGGGTCCACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCAGGTCCACAGGCCAGGCCACAGGAAGGGCAGGGGGACGCC
GGGGCCACAGCAGAGGCCGAGGAAGGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGGCTCCCTG
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCCTCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCTTAAAGAAGGTGGGATGGTGGCAATTTCTTGCCAGATTTTAGTCT
GCCCCGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGCACTGCTACCTACCTGTCTGCCC
5 GGGAGACAGGGAAAGCACCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCT
GCTCCAAATCACCACTTCTCTGGGGTTTTCCAAAGCATTTAAACAAGGGTGTACAGGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGCCCCCTGCTCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTGGTAAAAGGAAATGGTGAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAAACCCGAACACAGGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCTGGGGCCCTGCTGGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTTGTTCCTCCAGAGCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTCTTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCATTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCCTGTAGTCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAGTATCAGCATTCCAAAACCATAGTGACAGGTGTTTTTTATTC
TGTCCTTCGATAATTTTACTGGTGCTGTGCTAGAGGCCGGAACCTGGGGGTGCCTTCCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAACTGGGGTCTGTGCTGTTCTGAGTTAAGTCCAGATC
25 TGGACTTTGCTCTTTCCAGAATGCTCCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGGCCAGCATGTCCCTG
TTGCAGCTCCCTCCCAAGGATGCCGGTCTCCTGTGCTCCCAAGTCCCTGCTTCCCTCTCAGACCTTACCTGGTC
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGCTCCAGCACCTCTTCGCCTCTCCAGGCACCTCT
GCAGTGCTGGCCATACCAGTCAGCTGTGAAGTGTCCACTGCTTATTTTGCTCCCATGAAATGTATTTTATAGGACAGGC
30 ACCCTGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAATGG
GTTCTCTTAAACACATTGCAAAGCCACAGAGGCTAGTGAGGATGGGTGGGCATCAGGTATCAGATGTGGGTCCAATG
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGCAGAGGTGGCTCTAAAGCTCAGCAGTGGAG
GCAGTGGTTCGCCATACTCAGGGTGAACACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAAGAAAAGTAAAAAGGAAAAGTGGTAAGATGGGAATTTCTTGTCAGATTTTAGTC
35 TCCCAAACCAAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACAATAGAACAAACGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTGTAATTTTTTTCTGAGAAAAC
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAAAACAC
ACCAGCAACAGAAATAAAACAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAACTCAGCTTGCCTGAGCCACAGTGAA
40 ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGGCGCCCTGTGAGGTCTGCACATTATCCTCTCACTTT

5 GTTCTCCTAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCCAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAACTGACAGCGAGACCCATCCCTCAAAGAAACGACGCGTGAAGTGTGAGG
10 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAGCAGCTTGTTTGTCATGGAAGTCCTCACAATGTCTGTGTCTTCCCAGTAATTCACCTTCTGAAGTGA
CCAGACATTATCAGGGTCTTATTTACCATTTCCAGTGTTCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCAGAACTTGTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT
GTTTAATGGCAGAAACGTTTATTTCATGTAGCAGTGTTCAAAGCTGGATGTAAAAGAACACACCCAGGAGCCTGCCG
15 TGAATGTCATGTGTGTTTCATCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCCTCCATCCTGCCCCCTCTGGAGACACCATGTGTGCCACGTGCACTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCAACTCAGTGTCTCCCAAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCCGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTCTGGTG
15 GAGTTTGGTTCATGCAGAGTCTGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGACCCAGCCCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTCAGGAGGGGGCTGGGTGT
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCGT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACACGTGAGCACATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA
TGATGTGTGTTCTGTGCACAGTCGTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGCTAGGTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
25 GGGGTCCAGGCCTTGGTGGGCTGAGGCTCTGAAGTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACTCCCTCTCTGTGGGCATTTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGCAGGCAGATGACACAGAGTCTTGAATCGCCAGGTGGTTCGAGCTG
CCGGGTGAGGGCCAGGCCGATTCTCACTGGGAAGAGGGATAGTTTCTTGTCAAATGTTCTCTTTCTTGTTCATCTGA
30 ATGGATGATAAAGCAAAAAGTAAAACTTAAATCCCAGAGAGGTTTCTACCGTTTCTCACTCTTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT
TGAACTCCTGGGTTTATAGGGCAAGGAATGTCTTACGTTTTTCAGTGGTGTCTGCTGCCTGTGCACAGTTCGTGTCGCGTG
GCTCTGTGCAAAGCAGCTGTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT
GCACTGGCCGTGGGACGTCTAGGAGCCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGCTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGCTCCCCACAGCCCGGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGCTGGTGCAGGGGGTCCCCCTGCCAAGAATCG
40 ACAACTTTATCAGAGGGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTTGTGTTACCCAGGGCCG
AGGCTGCGCGAATTACCGTGACACTTGATGTGAAATGAGGTCGTCTATCGTGGAACCCAGCAAGGGCTCACGGGA
GAGTTTTCATTACAAGTGTACCATGAAAATGGTTTTTAACCCGAGTGTGCGCCTTCATGCTCTGGCAGGGAGGGC
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTCTGCGCTTCTCTCTGCGCTCAAATCTTCCCTCGTTTGATCTCCCTGACGCGTGCCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGGAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGTGCGGGCCTCCTTGGGCCATGATGAGGTGAGAGGAGTTTTC
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTC
10 ATTTCCCCACCAGGCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC
CCATGTGGGGACCCTTGGGTAGTGTCTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGCGCCGTGTGCGCGCCTCCACAGGGCTTGGGGTGGACGCCCCGACCTTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCATGCTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGGCGGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCTGGCTTCCGTTGTTGCTAAATGGGAAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGTGTGGCTTGACTGGTGTGATCTCAGGTATTCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGGCTCAGGCAGAGGCTCAGACCAGGTAC
20 ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC
TATGAAAAATAAAAAAATAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCAGGAGGTGGAAGCTGCAGTGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGAAAGAAACATTTAGTAGGAACCTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTGAGTGTGAGATGAGATGAGGTGCTCACACCATCACCCAGACCCAGGTTTATG
CACCAAGGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGGTTGTCTGACGAAGGGCAG
GATTCATGATAAGTACCTGTGTTACACAAGGAACAATGGATAAACTGGAACCTTAGAGGCCTTCCCGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGCAGTGTACCTGTGCACACACAGACACGCAGCTACTCGCACACACAAGCACACACACAGACATGCATGC
ATCGTGTGTGTGCACCTGTGCCATGAGGAAACCATGCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCCCAT
GCCCCACCCACGAGACCGTCTGATTAGGAGGCCTTTCCTCTGACGCTGTCCGCCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCCTGCTGGTGTAGTGTGTGAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCCCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCTGGGGAA
GCGCTGGGGGGCCTGGTCTCTCTGTTTGGCCCATGGTGGGATTGGGGGGCCTGGCCTCTCCTGTTTGGCCTGTGGTGG
GATTGGGGTGTCTCCCGTCCATGGCACTTAGGGCCCTTGTGCAAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCGTATACCACGACAGAGCCCCGCGCGTCTCTGCTTCCAGTCACCG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGCTCTGAAATTCAAGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCGACTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCGGGCCCCACCTGCCCAGGGGTCACTCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGCTGCTGAAGTGCAGACGCCCCCGGCTGACCTGGGGGCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCAGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCAGAGCCTCTTCCCTGGCTGTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCCGCCCCGCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCCGTTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACCAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCCGCCACCCACAGCTCCTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

10

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCGGGCCCAGTGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCGGGCCCAGT
GCCTGGCTGGGCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCCCAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGCGGCGGGGATGATG
GAGGGCCTGGCCAGGTGGCAGGGATGATGGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGCTGGTCTGGGTGG
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGGCCATCAGCTTTCATGGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCCAACTCAGGTTGAAAGTCACATTCGCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGTGGGTAGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTGTGCTCTCTC
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCCTCTTATCATCTCCAGTCTCATCTGTCTTCTCTTATCTCCAGT
CTCATCTGTCTATCCTTACCATCTCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCTCAGGCAGAAGGAACTGGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGAGTCTTGGGGTGAAGAAACAGCCCCTCCTCAGAAGTTGGCTTGGGCCACACGAAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGTAGTGCAGCCCGACGTGCCTGGTGTGCGGGTGGGGGCTTATGGCCACTGGATATG
GCGTCATTTATTGTGCTGCTTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCCAAGTCCACAGACTG
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCCGGGCGCGTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCAGGAGCAGAGGCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTTCAGTATCCATCTTCTACAAAGCTCCAGATTCTGTCTTCTCCGGGTGTTTTTGTGTAATTTTACTCAGGATTACT
TATATTTTTTGTCTAAAGTATTAGACCCCTAAAAAAGGTATTGTCTTTGATATGGCTTAACCTACTAAGCACCTACTTTAT
TTGTCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCAGGTGTAGTGCAGTGGCAC
AGTCATGGCTCGCTGTAGCCGCAAAACCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTGGGATTACAG
GTGTGAGCCACTGCCCTTGCTGGCACTTTTAAAAACCACTATGTAAGGTCAAGTCCAGTGGCTTCCACACCTGTCTATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAGACCAGCATGGGTAACATAGGGAGACCCC

20

25

30

35

40

ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCCTGTCTCAAAAAAAAAAAAAAAAAAAGAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG
5 GAAGAAGGAAGAAAGAAGGAGAAGGAGGCCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA
AAGTTTTAAAGGGAAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGACCTGTGTAACCGTC
GATGTTGGTGCCAGGTGCCCACCTGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
10 GCACTTGTGGCAGGCACAATTACAGCCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTCAACCG
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG
TGGGCCTGATATGGCCACAAGGGTCCCTAGAAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCTGAGGGCACACGGCCCTGCCACGCTCGATTTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTCAAGCCACTAAGCTGCAGTGATTTCGTACAGCAGCAAATGGAATAG
15 CAGTACAGGGAAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenght. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmen aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA, TTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_n/Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCAGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

20

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

25

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

30

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wasssergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄ pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2×10^5 -HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV
15 β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Fletcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- 20 Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 25 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 30 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol. 14:4894.
- 35 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- Harley, C.B., Fletcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

- Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.
- 5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.
- Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.
- 10 Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.
- 15 Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen
Telomerase-Untereinheit.
5
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser
Sequenzen handelt.
10
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
15
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß
es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide
oder Proteine kodieren.
20
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
25
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-
toren gemäß einem der Ansprüche 4 bis 6.
30

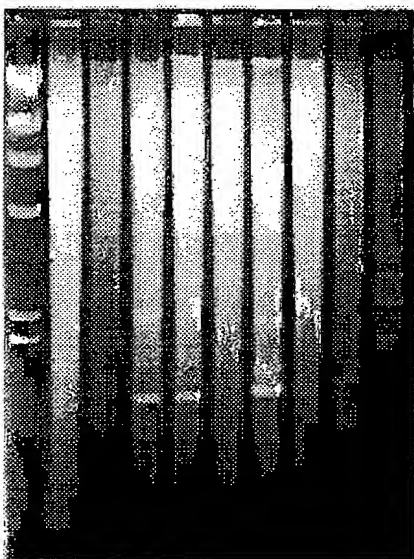
9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screenet.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reporter-genkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

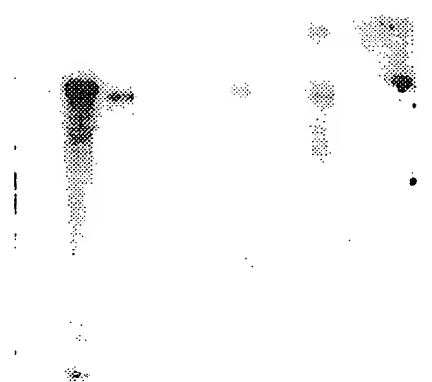


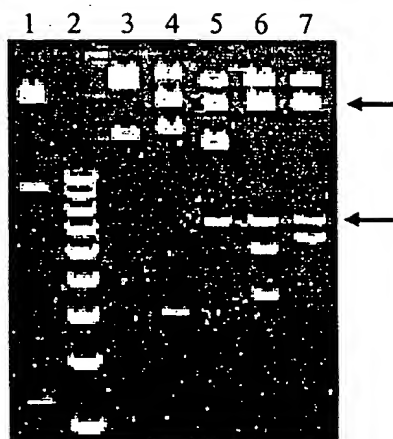
Fig. 2

Fig. 3

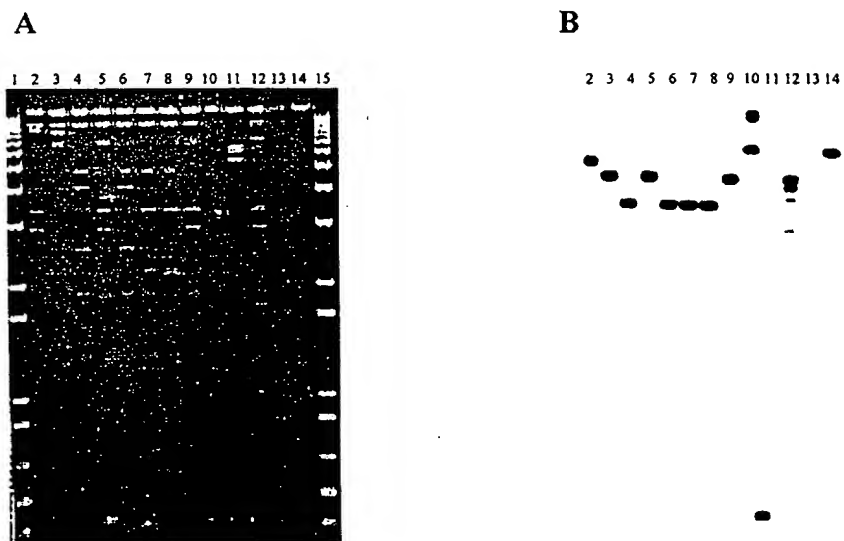


Fig. 4

GAGCTCTGAA	CCGTGGAAAC	GAACATGACC	CTTGCCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	70
GTGGTGTGCA	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCATTATT	140
CATCTTCACC	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAACTCAG	210
TACAAACACC	ACTCTTTTAC	TAGGCCACCA	GAGCAGGGGC	CACACCCCTG	ATATATTAAG	AGTCCAGGAG	280
AGATGAGGCT	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTCC	TCTAGACTAG	350
TAGACCCTGG	CAGGCACTCC	CCCAAATTCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	420
GAGACTCAGC	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCCT	CGCCCTCCAG	GCCTCAGCTT	490
CTCCAGCAGC	TTCCTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CACTGTGTCT	560
TGTCTCAGCG	ACGTAGCTCG	CACGGTTTCT	CCTCAGATGG	GGTGCTCTGC	TCCTTCCCCA	ACACTCACAT	630
GCGTTGAAGG	GAGGAGATTG	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	700
GATGCAGGTT	CCTGGCGTCC	GGCTGCACGC	TEACCTCCAT	TTCCAGGCGC	TCCCGCTCTC	CTGTTCATCTG	770
CCGGGGCCTG	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCTTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	840
CTAGGGTCTC	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GCGGCTCTTG	GGAAATGCAA	910
CATTTGGGTG	TGAAAGTAGG	AGTGCTGTTC	CTCACCTAGG	TCCACGGGCA	CAGGCTGGG	GATGGAGCCC	980
CCGCCAGGGA	CCCGCCCTTC	TCTGCCACGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	1050
TTCCACAAGC	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	1120
CCTGGGAATT	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGG	CTGTTTTATT	1190
TTAATAGCTA	CAAAAGCAGG	AAATCCCTGC	TAAATGTGCC	TTTAACAAAC	TGGTTAAACA	AACGGGTCCA	1260
TCCGCACGGT	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	1330
GAATTACGCT	CAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTC	1400
ACCCCATGGC	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	1470
CCTTTTACTA	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	1540
TGGGGATGGG	GGAAACCCGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	1610
AGAGATGCCC	ACGTCTGTAT	TCCCCCAAAC	CTGTGGCAGC	AACCCGCCCG	GCCCCAGGGC	CTTTGCAGGT	1680
GTGATCTCCG	TGAGGACCC	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	1750
GTTTCTGGGA	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCAGTCTGA	1820
GGCTGAAAGG	GAGGGGAGGG	CCTCGAGCCC	AGGCCTGCCA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	1890
GGGACCTCC	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	1960
TCCGGCCTCC	GTGCCATAGG	AGGGCAGCTG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTGCAGAA	2030
GCAACAGGAA	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAGGA	2100
CAGGGCTGAA	GTGCCTCCCG	GCAAGGGCAG	GCGAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	2170
TACTTACTTT	CTGAGACAGA	GTTATGCTCT	TGTTGCCCG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	2240
CTGCAACCTC	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTCAGGC	2310
GTGCACCAAC	ACACCCGGCT	AATTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	2380
TCTCAAAATC	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAAGT	CTGGGATTAC	AGGCATGAGC	2450
CACCTGCACCT	GGCCTATTTA	ACCATTTTAA	AACCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	2520
CATGGAGTTC	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAAATC	TTCTAGACT	2590
GGGGATACAC	CGTCTCTTGA	CATATTCCCA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCAGTGCAG	2660
GGGCAGCTGG	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	2730
ATCAGGGGCG	AAGTGTGGAC	ACTGTCCTGA	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	2800
AAGTCCATCC	CTCTACTCT	ACTGGGATTG	AGCCCCCTTC	CTATCCCCCC	CCAGGGGCAG	AGGAGTTCCT	2870
CTCCTCTCTG	TGGAGGAAGG	AATGATACTT	TGTTATTTT	CACCTGCTGT	ACTGAATCCA	CTGTTTCAAT	2940
TGTTGGTTTG	TTTGTTTTGT	TTTGAGAGGC	GTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	3010
CGCGATCTTG	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCCG	CTCCCATTTG	3080
GACTGGATTG	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GGGGGTGGGT	3150
GGGGTTCACC	ATGTTGGCCA	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCTCCT	3220
AAAGTGCTGG	GATTACAGGT	GTGAGCCACC	ATGCCACAGT	CAGAAATTTAC	TCTGTTTAGA	AACATCTGGG	3290
TCTGAGSTAG	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAAAT	TTTTTATTGT	3360
TGTTAGAACCA	CTCTTCATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	3430
TGCACCCATA	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	3500
CCATGCACAT	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTTCT	TCTCTTCCCT	CTTTTAAAAAT	3570
TGTGTTTTCT	ATGTTGGCTT	CTCTGCAGAG	AACCAGTGTA	AGCTACAAC	TAACCTTTGT	TGGAACAAAT	3640
TTTCCAAACC	GCCCCCTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	3710
ATCCTAAGG	GGATTCTTAG	AAGAGCGACC	TGTAATCTTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	3780
GAGCGTGACA	GCCCAGGGAG	GGTGCAGGCG	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	3850
GGCAGTTTCT	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTCACTGT	TTGCCGACCT	3920
CAGCTACAGC	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	3990
AACCCGGAGT	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTGTGTC	CGGGGCCCCA	GGTCTGGAGG	4060
GGACCACTGG	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGGCCTCCT	AGCTCTGCAG	TCCGAGGCTT	4130
GGAGCCAGGT	CCCTGGACCC	CGAGGCTGCC	CTCCACCTTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	4200
CATCTGCCAG	ACAGAGTGCC	GGGGCCAGG	GTCAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTCCCGC	4270
GCCAGCAGGA	GCGCTGGCT	CCATTTCCCA	CCCTTTCTCG	ACGGGACCGC	CCCGGTGGGT	GATTACAGAA	4340
TTTGGGGTGG	TGTGCTCATG	GTGGGGACCC	CTCGCCGCTT	GAGAACCTGC	AAAGAGAAAT	GACGGGCCCTG	4410
TGTCAAGGAG	CCCAAGTCCG	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCCGCCAG	4480
GGAGCAATGC	GTCTCTGGGT	TCGTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCCCTT	ACGTCCGGCA	4550
TTCTGTGGTG	CCGGAGCCCG	ACGCCCCCGG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	4620
GCGGCCAAG	GGTGCCTGCA	CGACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCCT	CGGGTTACCC	4690

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGCGGCCAG ACCCCCGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGAAGTG 4900
GGACCCGGGC ACCCGTCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGGCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126
```

Fig. 5

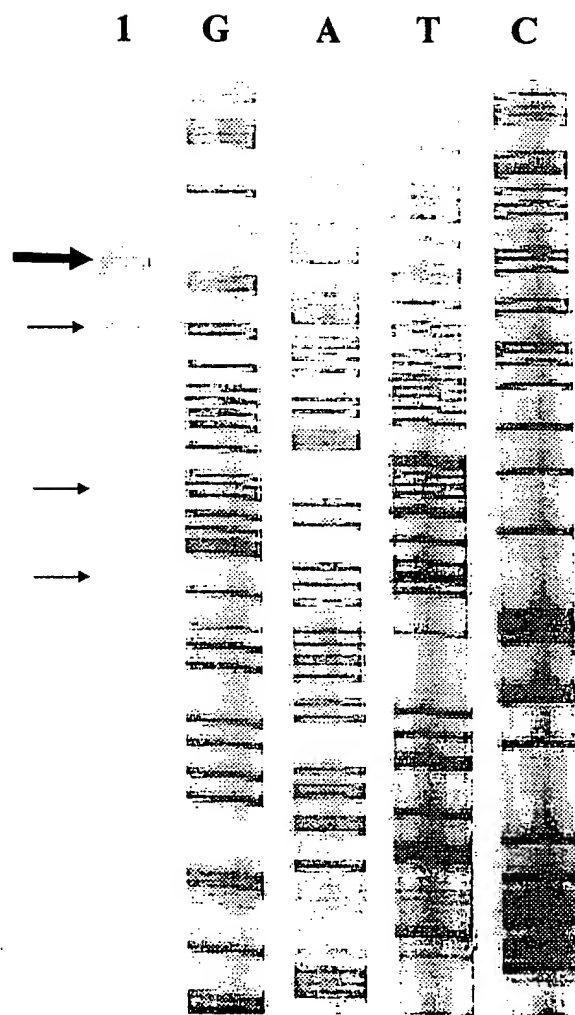


Fig. 6

GTTTCAGGCA	GCGCTGCGTC	CTGCTGCGCA	CGTGGGAAGC	CCTGGCCCCG	GCCACCCCCG	CGATGCCGCG	70
CGCTCCCCGC	TGCCGAGCCG	TGCGCTCCCT	GCTGCGCAGC	CACTACC CGC	AGGTGCTGCC	GCTGGCCACG	140
TTCGTGCGGC	GCCTGGGGCC	CCAGGGCTGG	CGGCTGGTGC	AGCGCGGGGA	CCCGCGGGCT	TTCCGCGCGC	210
TGGTGGCCCA	GTGCTGGTGC	TGCGTGCCCT	GGGACGCACG	GCCGCCCCCC	GCCGCCCCCT	CCTTCCGCCA	280
GGTGTCTGTC	CTGAAGGAGC	TGGTGGCCCG	AGTGTGTCAG	AGGCTGTGCG	AGCGCGGGCG	GAAGAAGCTG	350
CTGGCCTTCG	GCTTCGCGCT	GCTGGACGGG	GCCCGCGGGG	GCCCCCCCCG	GGCCTTCACC	ACCAGCGTGC	420
GCAGTCTGTC	GCCCAACACG	GTGACCGACG	CACTGCCGGG	GAGCGGGGGG	TGGGGGCTGC	TGCTGCGCCG	490
CGTGGCGGAC	GACGTGCTGG	TTCACCTGCT	GGCAGCGTGC	GCGCTCTTTG	TGCTGGTGGC	TCCGAGCTGC	560
GCCTACCAGG	TGTGCGGGCC	GCCGCTGTAC	CAGCTCGGCG	CTGCCACTCA	GGCCCGGGCC	CCGCCACACG	630
CTAGTGGACC	CCGAAGGCGT	CTGGGATGCG	AACGGGGCTG	GAACCATAGC	GTGAGGGAGG	CCGGGGTCCC	700
CCTGGGCCCTG	CCAGCCCCGG	GTGCGAGGAG	GCGCGGGGGC	AGTGCCAGCC	GAAGTCTGCC	GTTGCCCAAG	770
AGGCCCCAGG	GTGGCGCTGC	CCCTGAGCCG	GAGCGGACGC	CCGTTGGGCA	GGGGTCTCTG	GCCACCCCGG	840
GCAGGACGCG	TGACCCAGT	GACCGTGGTT	TCTGTGTGGT	GTCACCTGCC	AGACCCCGCC	AAGAGGCCAC	910
CTCTTTGGAG	GGTGGCTCT	CTGGCACGCG	CCACTCCAC	CCATCCGTGG	GCCGCCAGCA	CCACGCGGGC	980
CCCCCATCCA	CATCGCGGCC	ACCACGTCCC	TGGGACACGC	CTTGTCCTCC	GGTGTACGCC	GAGACCAAGC	1050
ACTTCTCTA	CTCTCAGGC	GACAAGGAGC	AGCTGCGGCC	CTCCTTCCTA	CTCAGCTCTC	TGAGGCCCAG	1120
CCTGACTGCG	GCTCGGAGGC	TGCTGGAGAC	CATCTTTCTG	GGTTCAGGCG	CCTGGATGCC	AGGGACTCCC	1190
CGCAGGTTGC	CCCGCTGCCC	CCAGCGCTAC	TGGCAATGTC	GGCCCTGTT	TCTGGAGCTG	CTTGGGAACC	1260
ACGCCGACGT	CCCTACGGG	GTGCTCCTCA	AGACGCACTG	CCCGCTGCGA	GCTGCGGTCA	CCCGACGACG	1330
CGGTGTCTGT	GCCCGGGAGA	AGCCCCAGGG	CTCTGTGGCG	GCCCCGAGG	AGGAGGACAC	AGACCCCTGT	1400
CGCCTGGTGC	AGCTGTCCG	CCAGCACAGC	AGCCCCCTGG	AGGTGTACGG	CTTCGTGCGG	GCCTGCCTGC	1470
GCCGGCTGTGT	GCCCCAGCG	CTCTGGGGCT	CACGGCACAA	GGAACGCCCG	TTCTCTAGGA	ACACCAAGAA	1540
GTTTATCTCC	CTGGGGAAGC	ATGCCAAGCT	CTCGCTGCAG	GAGCTGACGT	GGAAGATGAG	CGTGGCGGAC	1610
TGCGCTTGCG	TGCGCAGGAG	CCCAGGGGTT	GGCTGTGTTC	CGGCCGCGA	GCACCGTCTG	CGTGAGGAGA	1680
TCCTGGCCAA	GTTCCTGCAC	TGGCTGATGA	GTGTGTACGT	CGTCGAGCTG	CTCAGGTCTT	TCTTTTATGT	1750
CACGGAGACC	ACGTTTCAAA	AGAACAGGCT	CTTTTCTAC	CGGAAGAGTG	TCTGGAGCAA	GTTGCAAAAG	1820
ATTGGAATCA	GACAGCACTT	GAAGAGGGTG	CAGCTGCGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCAGCATC	1890
GGGAAGCCAG	GCCCCCCTG	CTGACGTCCA	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	1960
TGTGAACATG	GACTACGTCG	TGGGAGCCAG	AACGTTCCCG	AGAGAAAAGA	GGGCCGAGCG	TCTCACCTCG	2030
AGGGTGAAGG	CACGTGTCAG	CGTGCTCAAC	TACGAGCGGG	CGCGCGCGCC	CGGCCTCCTG	GCGGCCTCTG	2100
TGCTGGCCCT	GGACGATATC	CACAGGCGCT	GGCGCACCTT	CGTGTGCGT	GTGCGGGCCC	AGGACCCGCC	2170
GCCTGAGCTG	TACTTTGTCA	AGGTGGATGT	GACGGGCGCG	TACGACACCA	TCCCCCAGGA	CAGGCTCAGC	2240
GAGGTCATCG	CCAGCATCAT	CAAACCCAG	AACACGTACT	GCGTGGCTCG	GTATGCCGTG	GTCCAGAGAG	2310
CCGCCCATGG	GCACGTCCGC	AAGGCCTTCA	AGAGCCACGT	CTCTACCTTG	ACAGACCTCC	AGCCGTACAT	2380
GCGACAGTTC	GTGGCTCACC	TGCAGGAGAC	CAGCCCGCTG	AGGGATGCCG	TGCTCATCGA	GCAGAGCTCC	2450
TCCCTGAATG	AGGCCAGCAG	TGGCCTCTTC	GACGTCTTCC	TACGCTTCAT	GTGCCACCAC	GCCGTGCCCA	2520
TCAGGGGCCAA	GTCCTACGTC	CAGTGCCAGG	GGATCCCGCA	GGGCTCCATC	CTCTCCACGC	TGCTCTGCAG	2590
CCTGTGCTAC	GGCGACATGG	AGAACAAGCT	GTTTGGCGGG	ATTGCGCGGG	ACGGGCTGCT	CCTGCGTTTG	2660
GTGGATGATT	TCTTGTGGT	GACACCTCAC	CTCACCACAG	CGAAACCTT	CCTCAGGACC	CTGGTCCGAG	2730
GTGTCCCTGA	GTATGGCTGC	GTGGGAACT	TGCGGAAGAC	AGTGGTGAAC	TTCCCTGTAG	AAGACGAGGC	2800
CCTGGGTGCG	ACGGCTTTTG	TTCAGATGCC	GGCCACGGC	CTATTCCCTT	GGTGGCGGCT	GCTGTGGGAT	2870
ACCCGGACCC	TGGAGGTGCA	GAGCGACTAC	TCCAGCTATG	CCCGGACCTC	CATCAGAGCC	AGTCTCACCT	2940
TCAACCGCGG	CTTCAAGGCT	GGGAGGAACA	TGCGTCGCAA	ACTCTTTGGG	GTCTTGCGGC	TGAAGTGTCA	3010
CAGCCTGTTT	CTGGATTGTC	AGGTGAACAG	CCTCCAGACG	GTGTGCACCA	ACATCTACAA	GATCCTCCTG	3080
CTGCAGGCGT	ACAGGTTTCA	CGCATGTGTG	CTGCAGCTCC	CATTTTCATCA	GCAAGTTTGG	AAGAACCCCA	3150
CATTTTCTCT	GCGCGTCAATC	TCTGACACGG	CCTCCCTCTG	CTACTCCATC	CTGAAAGCCA	AGAACGACGG	3220
GATGTGCTG	GGGGCCAAAG	GCGCGCGCCG	CCCTCTGCCC	TCCGAGGGCT	TGCAGTGGCT	GTGCCACCAA	3290
GCATTCTCTG	TCAAGCTGAC	TCGACACCGT	GTCACCTACG	TGCCACTCCT	GGGGTCACTC	AGGACAGCCC	3360
AGACGACGCT	GAGTCGGAAG	CTCCCGGGGA	CGACGCTGAC	TGCCCCTGGG	GCCGCAGCCA	ACCCGGCACT	3430
GCCCTCAGAC	TTCAAGACCA	TCTTGACTG	ATGCCCACCC	GCCCACAGCC	AGGCCGAGAG	CAGACACCCG	3500
CAGCCCTGTC	ACGCCGGGCT	CTACGTCCCA	GGGAGGGAGG	GGCGGCCAC	ACCCAGGGCC	GCACCGCTGG	3570
GAGTCTGAGG	CCTGAGTGAG	TGTTTGGCCG	AGGCCTGCAT	GTCCGGCTGA	AGGCTGAGTG	TCCGGCTGAG	3640
GCCTGAGCGA	GTGTCCAGCC	AAGGGCTGAG	TGTCCAGCAC	ACCTGCGGTC	TTCACCTCCC	CACAGGCTGG	3710
CGCTCGGCTC	CACCCAGGG	CCAGCTTTTC	CTCACCAGGA	GCCCAGGCTT	CACTCCCCAC	ATAGGAATAG	3780
TCCATCCCCA	GATTCGCCAT	TGTTACCCCC	TGCGCCCTGCC	CTCCTTTGCC	TTCCACCCCC	ACCATCCAGG	3850
TGGAGACCTT	GAGAAGGACC	CTGGGAGCTC	TGGGAATTTG	GAGTGACCAA	AGGTGTGCCC	TGTACACAGG	3920
CGAGGACCTT	GCACCTGGAT	GGGGGTCCCT	GTGGGTCAAA	TTGGGGGGAG	GTGCTGTGGG	AGTAAATATC	3990
TGAATATATG	AGTTTTTCAG	TTTTGAAAAA	AAAAAAAAAA	AAAAAAAAAA	AA		4042

Fig. 7

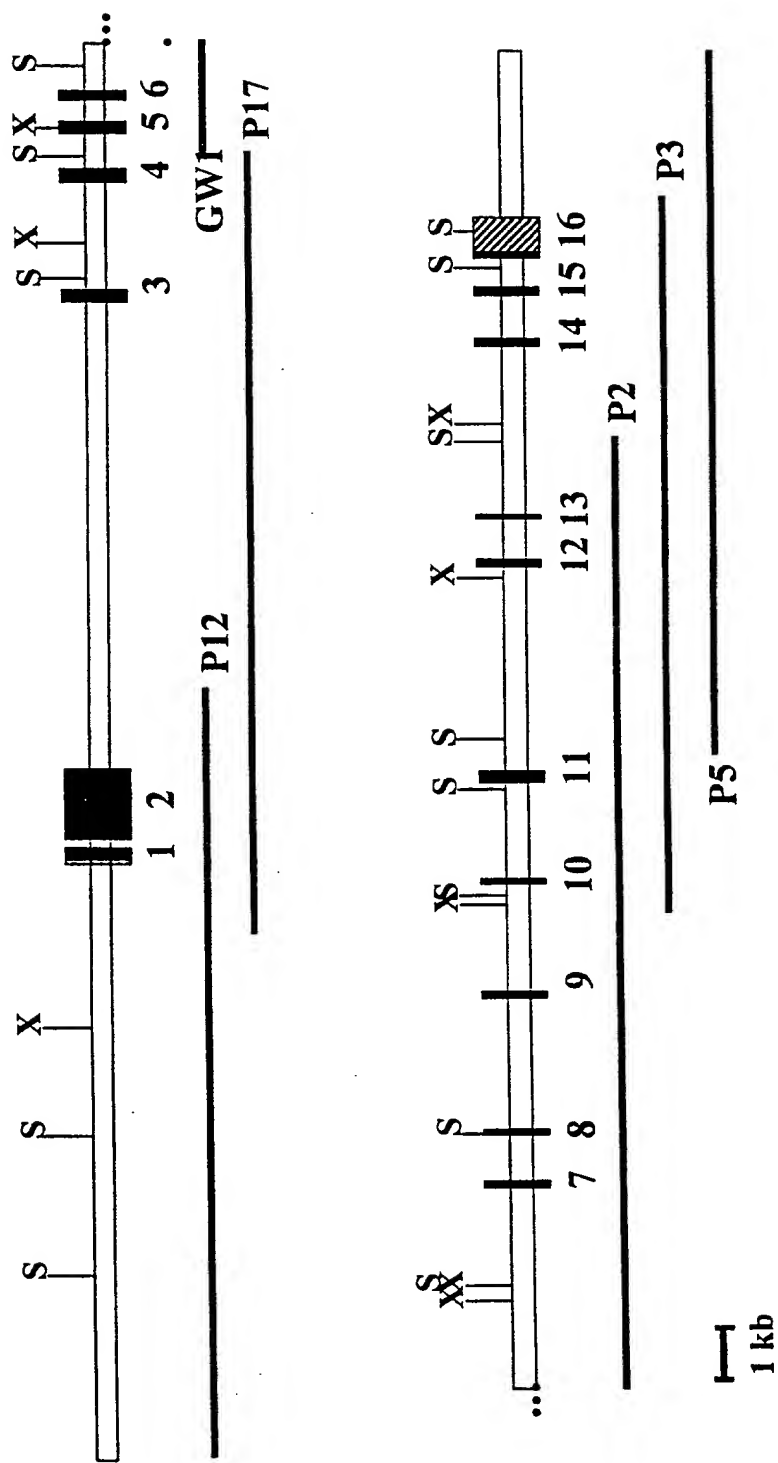


Fig. 8A

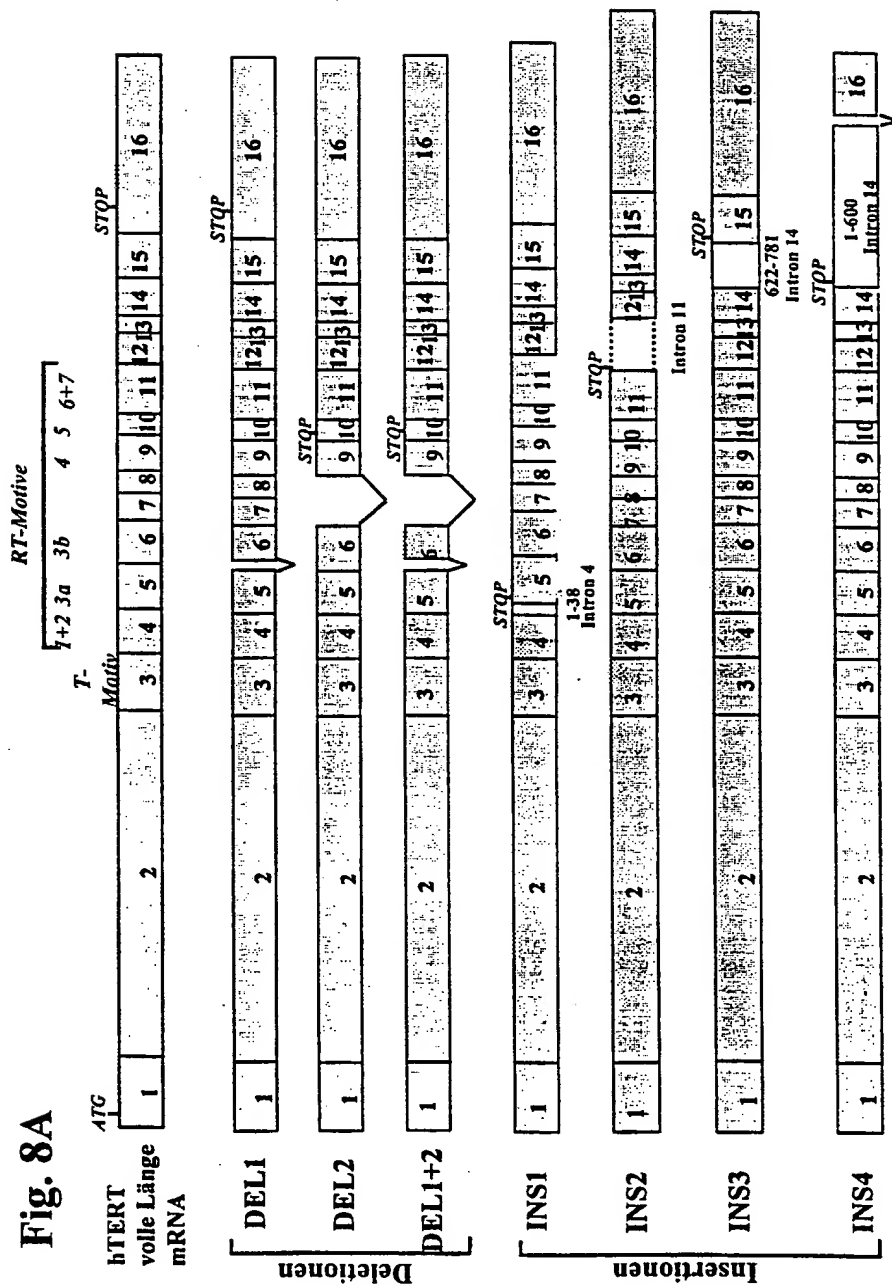


Fig. 8B

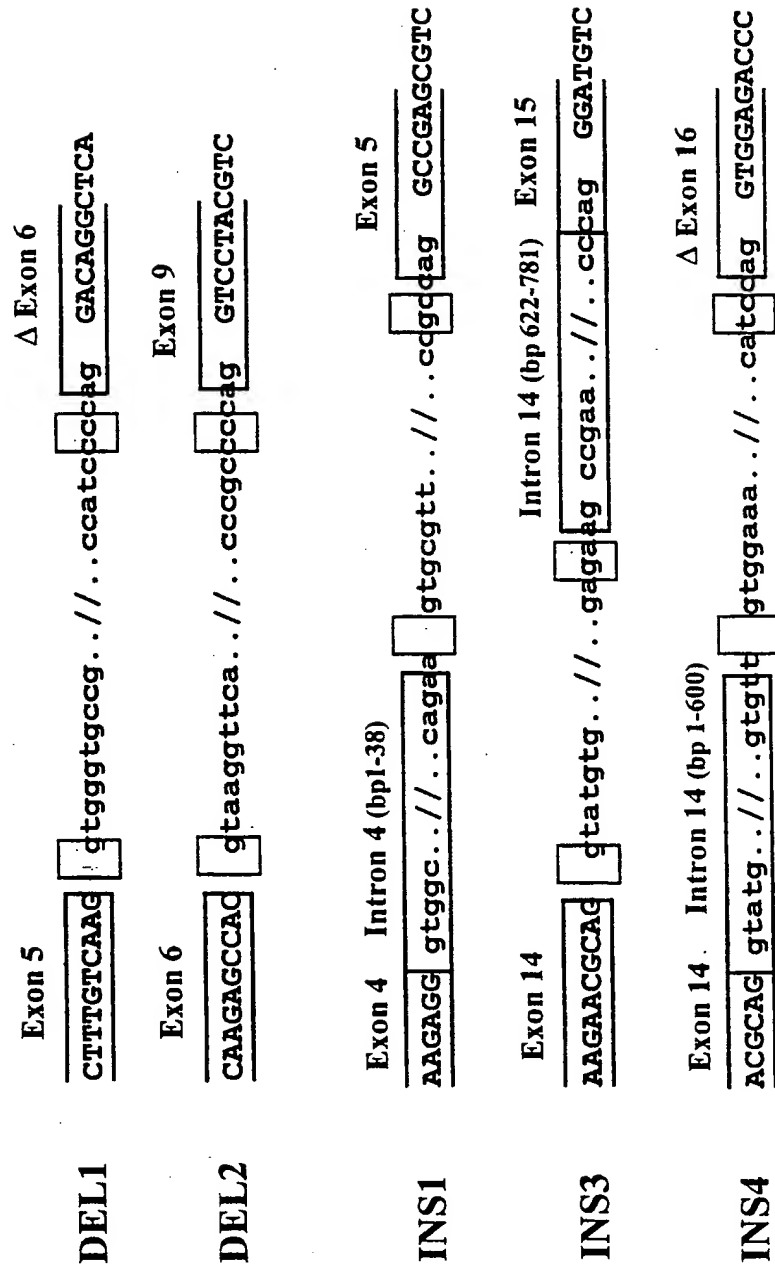


Fig. 9

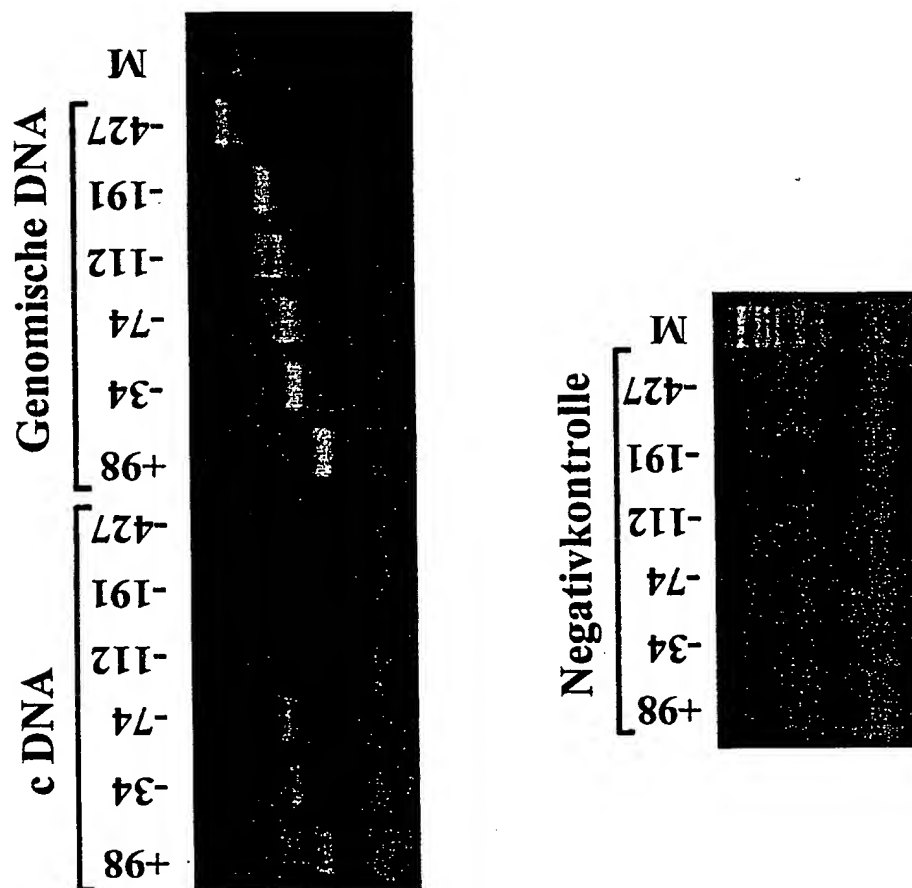


Fig. 10

ACTTGAGCCC AAGAGTTCAG GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCCT GTCTCAAAAA AAAAAAATAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGGG -11134
 ACAAACCCAG AAATCAACAA CAAGAGGAAT TTTGAAACT ATACAAACAC ATGAAAATTA AACAAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAATAAGGA AATTGAAAAA TTTATTAAAG CAAATGATAA -10994
 CGGAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG -10854
 GCCAAGGCGG GCAGATCGCC TGAGGTCAAG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG -10784
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAATAA AAGTAGAAAA ACTTAAAAAT ACACCTTAAT -10574
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAAATT GTAAAAAGAA AGAAATAATA -10504
 AAGATCAGAG CAGAAATAAA TGAACCTGAA AGATAACAAT ACAAAGATC AACAAAATTA AAAGTTGGTT -10434
 TTTTAAAAAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTCTAGA TGCATACAAAC -10224
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAA GACCAATAAC AATAATGGGA TTAAGCCAT -10154
 AATAAAAAAG TCCCTAGCAA AGGACCAACC GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084
 AAAGAAGAAT GAATTTCAAT CCTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACCTCAT -10014
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AACAAAAAAA -9944
 CAGAAAGAAA GAAAACTACA GGCCAAATAT CCTGATGAAT ACTGATACAA AAATCCTCAA CAAACACTA -9874
 GCAAACCAAA TTAACAACAA CCTTCGAAAG ATCATTCAAT GTGATCAAGT GGGATTATT CCAGGGATGG -9804
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAACACTA -9734
 TATGATTATT TCACCTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCTGATA AAAACCTCTA -9664
 AAAAAACGAG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TCGCATCCCA GCACCTCTGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT -9524
 CTACAAAAAA CTTTTTAAAA AAATTAGCCA GGCATGATGG CATATGCCTG TAGTCCAGC TAGTCTGGAG -9454
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAGTGAGC CATGAACATG TCACTGTACT -9384
 CCAGCCTAGA CAACGAGAACA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAACTGAA -9174
 AGCCTTTTCT CTAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAAA -9034
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964
 GCTGAAATTT GGTACAGCAG GATACAAAA CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAAAGAAA CAAAAAGCA GCTACAATAA AAATTAACA GCTAGGAATT AACCAAAGAA -8824
 GTGAAAGATC TCTACAATGA AAATAATAAA ATGTTGATAA AAGAAATTGA AGAGGGCACA AAAAAAGAAA -8754
 AGATATTCCA TGTTTATAGA TTGGAAGAA AAATCTGTT AAAATGTCCA TACTACCCAA AGCAATTTAC -8684
 AAATTCAGT CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614
 TTGTACAGAA CACAAAAGA CCCAGAATAG CCAAAGCTAT CCTGACCAA AAGAACAATA CTGGAAGCAT -8544
 CACATTACCT GACTTCAAT TATACTACAA AGCTATAGTA ACCCAAATA CATGGTACTG GCATAAAAC -8474
 AGATGAGACA TGGACCAAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTA -8404
 TTTTGGACAA AGGTGCAAG AACATACTTT GGGGAAAAGA TAATCTCTTC AATAAATGGT CTGGGAGGAA -8334
 CTGGATATCC ATATGCAAAA TAACAATACT AGAATCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264
 GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAATCTCCA -8194
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATTCCTG CAGGCACAGG CAACCAAAGC AAAACAGAC -8124
 AAATGGGATC ATATCAAGTT AAAAAAGCTT TGCCCAAGAA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA -7914
 CATTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704
 TTGCTACCAC TATGGAGAAC AGTTTGAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGTA TCAACAAGCT ATCTCCACT CCACATTTAC -7564
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCAT CAACAGACGA ATGGAAGGAG -7494
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424
 CAGCATGGGG GGCATGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
 CTCCCTTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG TTTCTAGAGG -7284
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAA GTGACAGATA CCCATTATAC CCTGATGTGA -7144
 TTATTACACA TTGTATGCC CTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCTTA CTATATTAAA -7074
 AATTAAAAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004
 GTGGATCACC TGAGGTCAGG AGTTTGAAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTAATAAGA -6934
 TACAAAAAT AGCCAGGCGT GGTGGCACAT ACCTGTAGTC CCACTACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTTGAAC CTGGGAGGCG GAGGTTCAG TGAGCCGAGA TCATGCCACT GCACCTGCAG CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAAA ACAAACAA AAAAAAGAG ATTAATAATTG TAATTTTAT GTACCGTATA -6724
 AATATATACT CTACTATATT AGAAGTTAAA AATTAAAAA ATTATAAAG GTAATTAACC ACTTAATCTA -6654
 AAATAAGAAT TGTATGTG GGGTTTCTG CTCTCAGAA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAAAC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCTCT TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA -6024
 AGGCGGCCAG CGGAATGCA AGGAGTCAGA AGCCTCTGGA TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCTTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCTTCTT CAAGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCTCTCT CCCTCTCTTG -5744
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCCTGGA GCTTCTCCGA GCCCGTGTG AGGACCCCTT -5674
 TGCAAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAC CTTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAATA ACATTCAGGA CTGCAGAAAT -5464
 CCAAAGCGGT TCCACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCTGCG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGTGCACCA TTGGGCAACG CGAAGGCGCG CACGCTGCGT -5184
 GTGACTCAGG ACCCCATACC GGCTTCTGCG CCCCACCCAC ACTAACCCAG GAAGTACGG AGCTCTGAAC -5114
 CCGTGGAAC GAACATGACC CTGCGCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGA -5044
 GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTATT CATCTTCACC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAACTCAG TACAAACACC -4904
 ACTCTTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTTACCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCTGG -4764
 CAGGCCTCC CCCAGATTCT AGGGCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGCC CACTGAGGC CAGCCCTGTC TCCACACCTT CCGCCTCCAG GCCTCAGCTT TCCAGCAGC -4624
 TTCCTAAACC CTGGGTGGGC CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554
 ACGTAGCTCG CACGGTTCTT CCTCAGATGG GGTGTCTGTC TCCTTCCCCA ACACCTCACAT GCGTTGAAG -4484
 GAGGAGATTG TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GTGCAGGTT -4414
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCATCTG CCGGGGCCCTG -4344
 CCGGTGTGTT CTTCTGTTTC TGTGCTCCTT TCCACGTCCA GCTGCGGTG TCTCTGCCG CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCCAG GGCCTCTTG GGAATGCAA CATTTGGGTG -4204
 TGAAGTAGG AGTGCTGTG CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCCGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT GCCCCACAGC CCTGGGAATT -3994

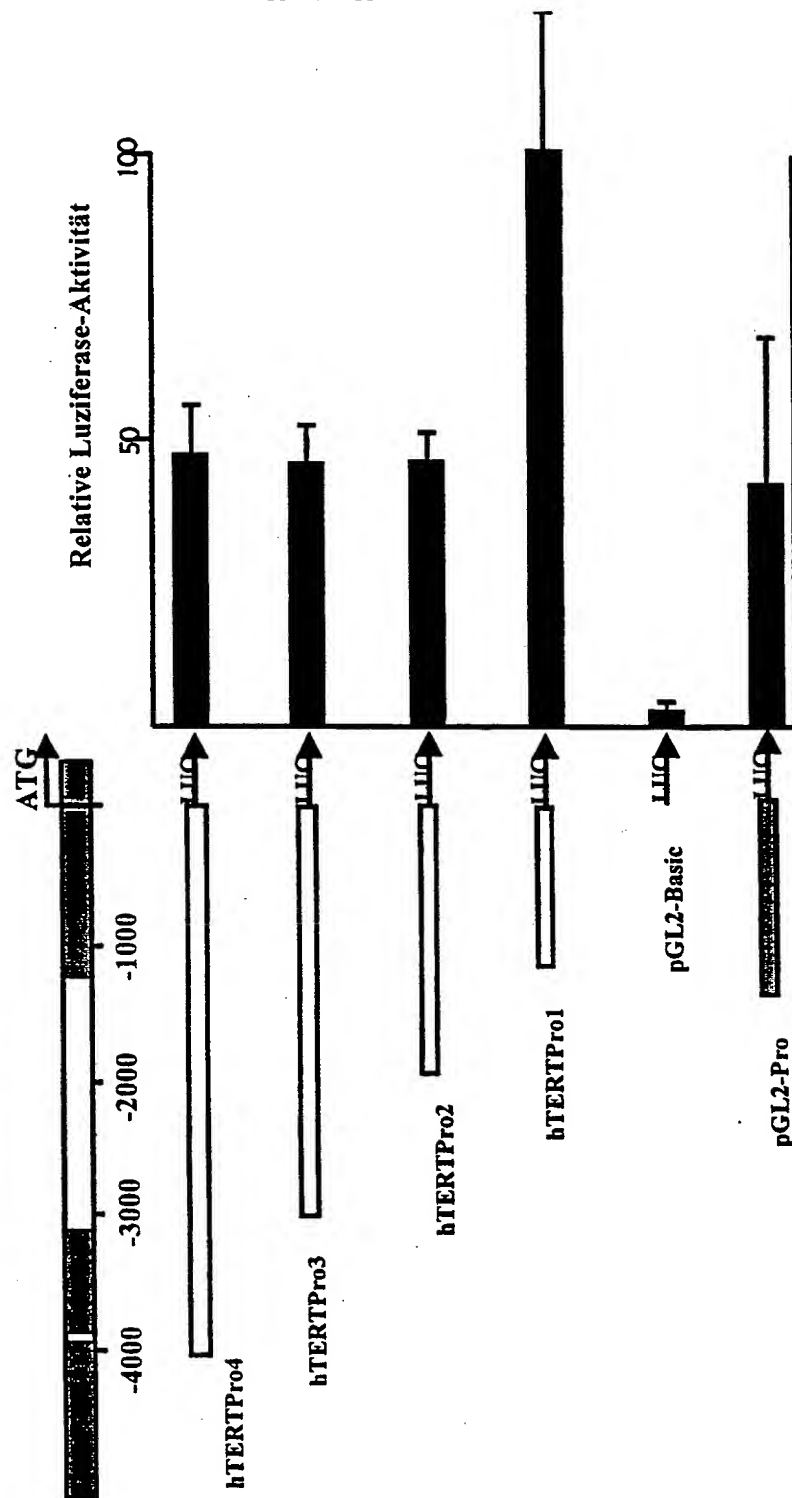
c-Myc
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGGTTAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAGAATTTC ACCCATGGC -3714
 AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644
 AAGCCAGTTT CTTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCC -3504
 ACGTCTGTAT TCCCCCAAAC CTGTGGACAG AACCCGCCC GCCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434
 TGAGGACCCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCCAAAA GTAATCCAGG GGTCTGCGGA -3364
 AGAGGGCGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCTCCAG AACCTGGAAA AAGCGGGGAA GGGACCCCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCCTCC -3154
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
 GTGCCTCCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCTGCG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTAGTAGA GATGGGCTTT CACCATGTTG GTCAGCTGA TCTCAAATC -2734
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTTCCCTT TACTCAGGA GTTACCTTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
 CGTCTCTTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCAG AGGAGTTCCT CTCACTCCTG -2244
 TGGAGGAAG AATGATACTT TGTATTTTCT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGGTTTG -2174
 TTTGTTTTGT TTTGAGAGGC GTTTCACTC TTGTTGCTCA GGTGGAGGG AGTGCAATGG CGCGATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTC AGTGATTCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA -2034
 CAGGCACCCG CACCATGCCC CAGCTAATTT TTTGATTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCAAC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
 GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTAC TCTGTTTGA AACATCTGGG TCTGAGGTAG -1824
 GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTACAAT-BoxGCCAA TSATAGAATT TTTTATTGT TGTAGAACA -1754
 CTCTTGATGT TTTACTGTGAT GATGACTAAG ACATCATCAG CTTTCAAAG ACACACTAAC TGCACCCATA -1684
 ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614
 GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTCCCT CTTTAAAAAT TGTGTTTTCT -1544
 ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTGT TGAACAAAT TTTCCAACC -1474
Spl
GCCGCTTTGC CCTAGTGGCA GAGACAATC ACAACACAG CCCTTTAAAA AGGCTTAGGG ATCCTAAGG -1404
 GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
 GCCAGGGAG GGTGCGAGGC CTGTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCTCTC GGCAGTTTCT -1264
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCACTGT TTGCCGACCT CAGCTACAGC -1194
 ATCCCTGCAA GGCCTCGGA GACCCAGAAG TTTCTCGCCC CTTAGATCC AAACCTGAGC AACCCGAGT -1124
 CTGGATTCTT GGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCAGTGG -1054
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGCCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
 GCCTGGACCC CGAGGCTGCC CTCCACCCTG TCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914
 ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844
 GCGCTGGCT CCATTCCAC-BoxCCCA CCCTTCTCG ACGGGASplCGC CCCGTTGGGT GATTAACAGA TTTGGGGTGG -774
 TTTGCTCATG GTGGGACCC CTCGCCGCT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCT GCGCGTCCAG GGAGCAATGC -634
 GTCCTCGGT AP-2TCCCCAG CCGCGTCTAC GCGCTCCGT CCTCCCTTC ACGTCCGCA TTCGTGGTGC -564
 CCGGAGCCCG ACGCCCGCG TCCGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
 GGTCGCCGA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCTCCCT CGGTTACCC CACAGCCTAG -424
 GCGATTGCA CCTCTCTCG CTGGGGCCCT CGTGGCGTC CCTGCACCCT GGGAGCGCA GCGGCSplGCGC -354
GCGGGGAAG CGCGGCCAG ACCCCCGGT SplCCGCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284
 CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC c-MycCCACGTGCG GAGGGACTGG GGACCCGGGC -214
 ACCCGTCTG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACSplCC GCCCGTCCC GACCCCTCCC -144
 GGGTCCCCG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTT TCCGCGGCC SplCGCCCTCTCC -74
c-Myc
TCGCGGCGG AGTTTCAGGC AGCGCTGCGT CTTGCTGCC ACGTGGGAAG CCCTGGCCCC GGCCACCCCC -4
 GCGATG

15 / 15

Fig.: 11



SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>
<141>

15 <160> 20

<170> PatentIn Vers. 2.0

<210> 1
<211> 5126
20 <212> DNA
<213> Homo sapiens

<400> 1

25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
gggtaaatgaa gtggtgtgca ggaaatggcc atgtaaatta cagcactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag taaaaacacc actcttttac taggcccaca 240
gagcacgggc cacacccctg atataatga agtcaggagg agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggt tgctgcttcc cgaggggccc atctgcccctg 420
gagactcagc ctggggtgcc acaactgaggc cagccctgtc tccacaccct ccgacctccag 480
gcctcagctt ctccagcagc ttcttaaac ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cagggttcc cctcacatgg 600
gggtgtctgc tccttccccca aactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720
ggctgcacgc tgacctccat ttccaggcgc tcccgtctc ctgtcatctg ccggggcctg 780
ccggtgtggt cttctgttct tgtgctcctt tccacgtcca gctgcgtgtg tctctgcccc 840
ctagggtctc ggggttttta taggcatagg acgggggcgt ggtggggccag ggcgctcttg 900
ggaaatgcaa catttgggtg tgaaagtagg agtgctctgc ctccactagg tccacgggca 960
40 caggcctggg gatggagccc ccgcccaggga cccgccttct tctgcccagc actttcctgc 1020
ccccctccc ctggaacaca gagtgcagc tccacaagc actaagcatc ctcttcccaa 1080
aagaccacgc attggcacc ctggacattt gccccacagc cctgggaatt cagtgacta 1140
cgcacatcat gtacacactc ccgtccacga ccgacccccg ctgttttatt ttaatagcta 1200
caaagcaggg aaatccctgc taaaatgtcc tttaacaaac tgggttaaaca aacgggtcca 1260
45 tccgcacggg ggacagttcc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc acccatggc aggggagtggt ttaggggggt taaggacggg 1440
ggggggcgga gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500
atggtattgg ctgattatg ggagactaac cataggggag tggggatggg ggaacccgga 1560
50 ggctgtgcca tctttgccat gcccgagtggt cctgggcagg ataatgctct agagatgccc 1620
acgtcctgat tccccaaac ctgtggacag aaccggcccg gccccagggc ctttgcaggg 1680
gtgatctccg tgaggacctg gaggctctgg atccttcggg actacctgca gggccgaaaa 1740
gtaatccagg ggttctggga agaggcgggc agggagggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggaggaggag cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctgaaa aagcggggaa gggaccctcc acggagcctg cagcagggaag 1920
gcacggctgg cccttagccc accagggcc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactcg cgctgcccct ctgagcatga gtgtgtgggg atttgcagaa gcaacaggaa 2040
acccatgcac tgtgaattca ggattattc aaaacaaagg ttacagaaa catccaagga 2100
cagggtgaa gtgcctcccg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
60 ttattttatt tacttacttt ctgagacaga gttatgctct tgttgcccag gctggagtgc 2220
agcggcatga tcttggtcga ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccgggct aattttgtat 2340
ttttagtaga gatgggcttt caccatggtg gtcaagctga tctcaaaatc ctgacctcag 2400
tgtatccgcc cacctcagcc tcccaaagtg ctgggattac aggcattgag cactgcacct 2460

2 / 18

5 ggccatttta accatttttaa aacttccctg ggctcaagtc acacccactg gtaaggagtt 2520
catggagttc aattttccctt ttactcagga gttacccctcc tttgatattt tctgtaattc 2580
ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
tcccatggga cccactgcag gggcagctgg gaggtgcag gcttcaggtc ccagtgggt 2700
tgccatctgc cagtagaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760
atctcaatgt ctcatgtgt gctgaaacat gtagaaatta aagtcacatcc ctccactct 2820
actgggattg agcccttcc ctatccccc ccaggggcag aggagttcct ctactcctg 2880
tggaggaagg aatgatactt tgttattttt cactgctggg actgaatcca ctgtttcatt 2940
tgttggtttg tttgttttgg tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000
10 agtgcaatgg cgcgatcttg gcttactgca gcctctgcct cccagggtca agtgattctc 3060
ttgcttccgc ctccatttg gctgggatta caggcaccgc ccaccatgcc cagctaattt 3120
cttgattttt tagtagagac ggggttgggt ggggttcacc atgttgcca ggctgggtctc 3180
gaactttcga cctcagatga tccacctgcc tctgcctcct aaagtgtgg gattacaggt 3240
gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
15 gaagctcacc ccactcaagt gttgtggtg ttaagccaa tgatagaatt tttttattgt 3360
tgtagaaca ctcttgatgt ttacactgt gatgactaag acatcatcag cttttcaaag 3420
acacactaac tgcaccata atactgggtg gtcttctggg tatcagcaat cttcattgaa 3480
tgccggaggc cgtttctcgc ccactgcacat ggtgttaatt actccagcat aatcttctgc 3540
ttccatttct tctcttccct cttttaaatt tgtgttttct atgttggtt ctctgcagag 3600
20 aaccagtgtg agctacaact taacttttgt tggaacaaat tttccaaacc gcccttttgc 3660
cctagtggca gagacaattc acaaacacag ccctttaaata aggcttaggg atcactaagg 3720
ggattttctag aagagcgacc tgtaatccta agtatattaca agacgaggct aacctccagc 3780
gagcgtgaca gccagggagc ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
aatttctctc ggcagtttct gaaagttaga aaggttacat ttaaggttgc gtttgttagc 3900
25 atttcagtgt ttgccgacct cagctacagc atccctgcaa ggcctcggga gaccagaag 3960
tttctcgccc ccttagatcc aaacttgagc aaccggagc ctggattcct gggaagtcct 4020
cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
tctactgctg ggctggaagt cgggcctcct agctctgcag tccgaggctt ggagccaggt 4140
gcctggacc cagaggtgcc ctccacctg tgccggcggg atgtgaccag atgttgacct 4200
30 catctgccag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
ccggtgcgc gccagcaggc gcgcctggct ccatttccca cccttctcgc acgggaccgc 4320
cccggtgggt gattaacaga tttgggtggg ttgtctcatg gtggggacc ctcgccgcct 4380
gagaaactgc aaagagaaat gacgggcctg tgtcaaggag cccaagtgcg ggggaagtgt 4440
tgcaaggagg cactccggga ggtcccgcgt gccctccag ggagcaatgc gtcctcgggt 4500
35 tgcgccag ccgcgtctac gcgcctcct cctcccttc acgtccgca ttcgtgggtg 4560
ccggagccc cgcgccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620
gcggccaaag ggtgcgcga cgcacctgt cccagggcct ccacatcatg gcccttccct 4680
cgggttacc cacagcctag gccgattcga cctctctccg ctggggccct cgctggcgtc 4740
cctgcacct gggagcgcga gcggcgcgcg ggccgggaag cgcggcccag acccccgggt 4800
40 ccgcccggag cagctgcgct gtcggggcca ggccgggctc ccagtggatt cgcgggcaca 4860
gagcccaagg accgcgtcc ccacgtggcg gagggactgg ggaccgggg acccgtctctg 4920
ccccttacc ttccagctcc gcctcctccg cgcggacccc gccccgtccc gaccctccc 4980
gggtccccg cccagccccc tccgggcct cccagccct ccccttctct tccgcggccc 5040
cgccctctcc tcggcgcgcg agtttcaggc agcgtctgct cctgctgcgc acgtgggaag 5100
45 ccctggcccc ggccaccccc gcgatg 5126

<210> 2
<211> 4042
<212> DNA
50 <213> Homo sapiens

<400> 2
gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
cgatgccgcg cgctccccgc tgcgagccg tgcgtccct gctgcgcagc cactaccgcg 120
55 aggtgctgcc gctggccacg ttcgtgcgpc gctggggcc ccagggtgg cggctggtgc 180
agcgcgggga cccggcggct ttcgcgcgc tgggtggcca gtgcctggtg tgcgtgccct 240
gggacgcagc gccgcccccc gccgccccct ctttcgcga ggtgtcctgc ctgaaggagc 300
tgggtggccg agtgctgcag aggtgtgctg agcgcggcgc gaagaacgtg ctggccttgc 360
gcttcgcgct gctggacggg gcccgccggg gccccccga gcccttcacc accagcgtgc 420
60 gcagctacct gcccaacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480
tgctgcgcgc cgtgggcgac gacgtgctgg ttacactgct ggcacgctgc gcctctttg 540
tgctggtggc tcccaagctgc gcctaccagg tgtgcgggccc gccgtgtac cagctcggcg 600
ctgccactca ggcccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660
aacgggcctg gaaccatagc gtcaggggag ccgggggtccc cctgggcctg ccagccccgg 720
65 gtgcgaggag gcgcgggggc agtgccagcc gaagtctgct gttgcccaag aggccaggc 780

5 gtggcgctgc ccctgagccg gacgcggacgc ccgttgggca ggggtcctgg gccaccccg 840
 gcaggacgcg tggaccgagt gaccgtgggt tctgtgtggt gtcacctgcc agaccgcgcg 900
 aagaagccac ctctttggag ggtgcgctct ctggcacgcg ccactccac ccatccgtgg 960
 gccgccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020
 ctgtccccc ggtgtacgcc gagaccaagc acttctcta ctctcaggc gacaaggagc 1080
 agctgcgggc ctcttctcta ctacgtctc tgaggccag cctgactggc gctcggaggc 1140
 tegtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcaggttgc 1200
 cccgcctgcc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260
 acgcgacgtg cccctacggg gtgtctctca agacgcactg cccgctgcga gctgcgggtca 1320
 10 ccccgacgac cgtgtctgt gcccgggaga agccccaggg ctctgtggcg gcccccgagg 1380
 agggaggacac agacccccgt cgcctgggtg agctgtctcg ccagcacagc agccccctggc 1440
 aggtgtacgg ctctcgtcgg gcctgcctgc gccggtcgtg cccccaggc ctctggggct 1500
 ccaggacaaa cgaacgcgcg ttctctagga acaccaagaa gtatcatctc ctggggaagc 1560
 atcggaagct ctctgtcag gagctgacgt ggaagatgag cgtgcgggac tgcgcttggc 1620
 15 tgcgcaggag cccagggggt ggctgtgttc cggccgcaga gcaccgtctg cgtgaggaga 1680
 tcttggcaca gttcctgcac tggctgatga gtgtgtacgt cgtcgagctg ctacgggtctt 1740
 tcttttatgt caccggagac acgtttcaaa agaacaggct ctttttctac cggaagagt 1800
 tctggagcaa gttgcaaac attggaatca gacagcactt gaagagggtg cagctgcggg 1860
 agctgtcggg agcagaggtc aggcagcatc gggaagccag gcccgccctg ctgacgtcca 1920
 20 gactccgctt catccccaa cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg aggggtgaagg 2040
 cactgttcag cgtgctcaac tacgagcggg cgcggcgccc cggcctcctg ggcgctctg 2100
 tgcctgggct ggacgatat cacagggcct ggcgacactt cgtgtctgct gtgcgggccc 2160
 aggaccgcgc gcctgagctg tactttgtca aggtggatgt gacgggcgcg tacgacacca 2220
 25 tccccaggga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgcgtcg gtatgccgtg gtccagaagg ccgcccagtg gcacgtccgc aaggccttca 2340
 agagccacgt ctctacctg acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
 tgcaggagac cagcccgctg agggatgccc tctgtcatga gcagagctcc tccctgaatg 2460
 agggcagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcgca 2520
 30 tcaggggcaa gtccctacgtc cagtgccagg ggatcccga gggctccatc ctctccacgc 2580
 tgcctgtcag cctgtgtctac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640
 acgggctgct cctgcgtttg tgggatgatt tcttgttggg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760
 35 tgcggaagac agtgggtgaa ttccctgtag aagacgaggg cctgggtggc acggcttttg 2820
 ttcatagatc ggccacagcg ctattcccc tggcgccct gctgtctgat acccggaacc 2880
 tggagggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcacct 2940
 tcaaccgcgg ctccaaggct gggagggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000
 tgaagtgtca cagcctgttt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060
 40 acatctacaa gatcctcctg ctgcaggcgt acaggtttca cgcagtgtgt ctgcagctcc 3120
 catttcatca gcaagtttgg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180
 cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgctg ggggccaagg 3240
 gcgcgcggcg cctcttgcgc tccgagggcg tgcagtggct gtgccacca gcatctctgc 3300
 tcaagctgac tcgacacgtg gtcacctacg tgccactcct ggggtcactc aggacagccc 3360
 45 agacgcagct gagtgcgaag ctccccggga cgacgtgac tgccctggag gccgcagcca 3420
 acccggcact gccctcagac ttcaagacca tcttggactg atggccacc gccacagcc 3480
 agggcgagag cagacaccag cagccctgtc acgcccggct ctacgtccca gggaggagg 3540
 ggcggccccc acccaggccc gcaccgctgg gactctgagg cctgagttag tgtttggcg 3600
 aggcctgcag gtccggctga aggtgagtg tccggctgag gcctgagcga gtgtccagc 3660
 50 aagggtgag tgtccagcac acctgcgtc ttacttccc cacaggctgg cgtcgggtc 3720
 caccacaggg ccagcttttc ctaccagga gcccggttcc cactccccc ataggaatag 3780
 tccatcccca gattcgccat tgttaccctc tgccttgc ctcttttgc ttccaccccc 3840
 accatccagg tggagaccct gagaaggacc ctgggagctc tgggaatttg gagtgaacca 3900
 aggtgtgccc tgtacacagg caggagacct gcacctggat gggggtccct gtgggtcaaa 3960
 55 ttggggggag gtgctgtggg agtaaaatac tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3
 <211> 11276
 <212> DNA
 60 <213> Homo sapiens

<400> 3
 acttgagccc aagagttcaa ggctacgggtg agccatgatt gcaacaccac acgccagcct 60
 tggtagaga atgagaccct gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120
 65 tcttctctgg ccacagtgga acaaaaccag aaatcaacaa caagagggaat tttgaaaact 180

	atacaaacac	atgaaaatta	aacaatatac	ttctgaatga	ccagtgagtc	aatgaagaaa	240
	ttaaaaagga	aattgaaaaa	tttattttaag	caaatgataa	cggaacata	acctctcaaa	300
	accacacggt	tacagcaaaa	gcagtgctaa	gaaggaaagt	tatagctata	agcagctaca	360
	tcaaaaaagt	agaaaagcca	ggcgagtggt	ctcatgcctg	taatccagc	actttgggag	420
5	gccaaagcgg	gcagatcgcc	tgaggtcagg	agttcgagac	cagcctgacc	aacacagaga	480
	aaccttctgc	ctactaaaaa	tacaaaatta	gctgggcatg	gtggcacatg	cctgtaatcc	540
	cagactactcg	ggaggctgag	gcagataaac	cgcttgaacc	caggagggtg	agggttgcgt	600
	gagccgggat	tgccgcatg	gactccagcc	tggttaacaa	gagtgaacc	ctgtctcaag	660
	aaaaaaaaa	aagtagaaaa	acttaaaaat	acaacctaat	gatgcacctt	aaagaactag	720
10	aaaagcaaga	gcaaaactaa	cctaaaattg	gtaaaaagaa	agaaataata	aagatcagag	780
	cagaataaaa	tgaaactgaa	agataaccaat	acaaaagatc	aacaaaatta	aaagttggtt	840
	ttttgaaaag	ataaacaaaa	ttgacaaacc	tttgcccaga	ctaagaaaaa	aggaaagaag	900
	acctaataaa	ataaagtcag	agatgaaaaa	agagacatta	caactgatac	cacagaatt	960
	caaaggatca	ctagaggcta	ctatgagcaa	ctgtacacta	ataaattgaa	aaacctagaa	1020
15	aaaataagata	aattcctaga	tgcatacaac	ctaccaagat	tgaacctga	agaaatccaa	1080
	agcccaaaaa	gaccaataac	aataatggga	ttaaagccat	aataaaaaagt	ctcctagcaa	1140
	agagaagccc	aggacccaat	ggcttccctg	ctggatttta	ccaatcattt	aaagaagaat	1200
	gaattccaat	cctactcaaa	ctattctgaa	aaatagagga	aagaatactt	ccaaactcat	1260
	tctacatggc	cagtattacc	ctgattccaa	aaccagacaa	aaacacatca	aaaacaaaca	1320
20	aacaaaaaaa	cagaagaaaa	gaaaactaca	ggccaatatc	cctgatgaat	actgatacaa	1380
	aaatcctcaa	caaaacacta	gcaaaccaaa	ttaaacacaa	ccttcgaaaag	atcattcatt	1440
	gtgatcaagt	gggatttatt	ccagggatgg	aaggatgggt	caacatatgc	aaatcaatca	1500
	atgtgataca	tcattcccaac	aaaatgaaat	acaaaaacta	tatgattatt	tcactttatg	1560
	cagaaaaagc	atttgataaa	attctgcacc	cttcatgata	aaaacctca	aaaaaccagg	1620
25	tatacaagaa	acatacaggc	caggcacagt	ggctcacacc	tgcatcccca	gcactctggg	1680
	aggccaaggt	gggatgattg	cttgggccc	ggagtttgag	actagccttg	gcaacaaaat	1740
	gagacctgtg	ctacaaaaaa	cttttttaaa	aaattagcca	ggcatgatgg	catatgcctg	1800
	tagtcccagc	tagtctggag	gctgaggtgg	gagaatcact	taagcctagg	aggctcaggc	1860
	tgagtgagc	catgaacatg	tcactgtact	ccagcctaga	caacagaaca	agaccccaact	1920
30	gaataaagaag	aaggagaagg	agaaggga	agggagggag	aagggaaggag	gaggagaagg	1980
	aggaggtgga	ggagaagtgg	aaggggaagg	ggaaagggaaa	gaggaagaag	aagaaacata	2040
	tttaacacata	ataaaagccc	tatatgacag	accgaggtag	tattatgagg	aaaaactgaa	2100
	agcctttcct	ctaagatctg	gaaaatgaca	agggcccact	ttcaccactg	tgattcaaca	2160
	tagtactaga	agtcctagct	agagcaatca	gataagagaa	agaaataaaa	ggcatccaaa	2220
35	ctggaaagga	agaagtcaaa	ttatcctgtt	tgcatgatgat	atgatcttat	atctggaaaa	2280
	gacttaagac	accactaaaa	aactattaga	gctgaaattt	ggtacagcag	gatacaaaat	2340
	caatgtacaa	aaatcagtag	tattttctata	ttccaacagc	aaacaatctg	aaaaagaaac	2400
	caaaaaagca	gctacaaaata	aaattaaaaca	gctaggaatt	aaccaaagaa	gtgaaagatc	2460
40	tctacaatga	aaactataaa	atgttgataa	aagaaattga	agagggcaca	aaaaaagaaa	2520
	agataattcca	tgttcataga	ttggaagaat	aaatactgtt	aaatgtcca	tactacccaa	2580
	agcaatttac	aaattcaatg	caatccctat	taaaatacta	atgacgttct	tcacagaat	2640
	agaagaaaca	attctaagat	ttgtacagaa	ccacaaaaga	cccagaatag	ccaaagctat	2700
	cctgaccaa	aagaacaaaa	ctggaagcat	cacattacct	gacttcaaat	tatactacaa	2760
	agctatagta	accacaaacta	catggtactg	gcataaaaa	agatgagaca	tggaaccagag	2820
45	gaacagaata	gagaatccag	aaacaaatcc	atgcatctac	agtgaactca	tttttgacaa	2880
	agggtccaa	aacatacttt	ggggaaaaaga	taatctcttc	aataaatggt	gctggaggaa	2940
	ctggatatcc	atatgcaaaa	taacaatact	agaactctgt	ctctcaccat	atacaaaagc	3000
	aaatcaaaat	ggatgaaagg	cttaaatcta	aaacctcaaa	ctttgcaact	actaaaagaa	3060
	aacaccggag	aaactctcca	ggacattgga	gtgggcaaa	acttcttgag	taattccctg	3120
50	caggcacagg	caaccaaagc	aaaaacagac	aaatgggac	atatcaagtt	aaaaagcttc	3180
	tgccagcaaa	aggaacaat	caacaaagag	aagagacaac	ccacagaatg	ggagaatata	3240
	tttgcaaaat	attcatctaa	caaggaatta	ataaccagta	tatataagga	gctcaaaacta	3300
	ctctataaga	aaaacaccta	ataagctgat	tttcaaaaat	aagcaaaaaga	tctgggtaga	3360
	cattttctcaa	aataagtcac	acaaatggca	aacaggcatc	tgaaaatgtg	ctcaacacca	3420
55	ctgatcatca	gagaaatgca	aatcaaaaact	actatgagag	atcatctcat	cccagttaaa	3480
	atggctttta	ttcaaaagac	aggcaataac	aaatgccagt	gaggatgtgg	ataaaaggaa	3540
	acccttgagc	actgttggtg	ggaatggaaa	ttgctaccac	tatggagaac	agtttgaaag	3600
	ttcctcaaaa	aactaaaaat	aaagctacca	tacagcaatc	ccattgctag	gtatatactc	3660
	caaaaaaggg	aatcagtgta	tcaacaagct	atctccactc	ccacatttac	tgagcagctg	3720
60	ttcatagcag	ccaagggttg	gaagcaacct	cagtgtccat	caacagacga	atggaaaaag	3780
	aaaatgtggt	gcacatacac	aatggagtac	tacgcagcca	taaaaaagaa	tgagatcctg	3840
	tcagttgcaa	cagcatgggg	ggcactggtc	agtatgttaa	gtgaaataag	ccaggcacag	3900
	aaagacaaac	ttttcatggt	ctcccttact	tgtgggagca	aaaattaaaa	caattgacat	3960
	agaaatagag	gagaatggtg	gttctagagg	ggtgggggac	agggtgacta	gagtcacaaa	4020
65	taattttatg	tatgttttaa	aataactaaa	agagtataat	tgggttggtt	gtaacacaaa	4080

5 gaaaggataa atgcttgaag gtgacagata ccccatTTac cctgatgtga ttattacaca 4140
 ttgtatgctt gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
 aattaaaaat ttaattggcca ggcacgggtg ctcagtccg taatcccagc actttgggag 4260
 gccgaggcgg gtggatcacc tgaggtcagg agtttgaaac cagtctggcc accatgatga 4320
 aacctgtctt ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgagc 4440
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
 acaaaaacaa aaaaaagaag attaaaattg taatttttat gtaccgtata aatatatact 4560
 ctactatatt agaagttaaa aattaaaaca attataaaag gtaattaaac acttaactta 4620
 10 aaataagaac aatgtatgtg ggggttctag cttctgaaga agtaaaagt ttggtccaga 4680
 tggcagaaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctg 4740
 taagtacttt aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800
 cctaaaacaa ctgctaataa tgggtgaaag taatctctat taattaccaa taattacaga 4860
 tatctctaaa atcagagctgc agaattggca cgtctgatca caccgtctc tcattcacgg 4920
 15 tgcttttttt ctgtgtgtgt tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980
 atctgtatga atcctgaaac gaaaaatggg ggtgatttcc tccagaagaa ttagagtacc 5040
 tggcagggaag caggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
 aagaccaggc tgcaaggcag aggcctgatg acccgaggac aggaagctc ggatgggaag 5160
 gggcgatgag aagcctgcct cgttgggtgag cagcgcatga agtgccctta ttacgcttt 5220
 20 gcaaagattg ctctggatac catctggaag aggcggccag cgggaatgca aggagtcaga 5280
 agcctcctgc tcaaaaccag gccagcagct atggcgccca cccgggcgtg tggcaggagg 5340
 agaggagtca aggcacctcg aagtatggct taaatctttt ttccacctga agcagtgacc 5400
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatggaa 5460
 25 gcaccttctt caagggaaaa ccagacgccc gctctgcggg cattacctc ttctctctct 5520
 cctctctctg cctcgcgggt ttctgatcgg gacagagtga ccccggtgga gcttctccga 5580
 gccggtgctg aggcacctct tgcaaggggc tccacagacc cccgccttgg agagaggagt 5640
 ctgagcctgg cttataaaca aactgggatg tggctggggg cggacagcga cggcgggatt 5700
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760
 30 tctttaatatt ttcttaaaat tcatcaata acattcagga ctgcagaaat ccaaaggcgt 5820
 aaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttcagaggg 5880
 atttttctgc ctaagtactt ttatttggtt ttcataaggt ggcttagggg gcaagggaag 5940
 gtacacgagg agagcgcttg gcggcagggc tatgagcagc gcagggccac cggggagaga 6000
 gtccccggcc tgggaggctg acagcaggac cactgaccgt cctccctggg agctgccaca 6060
 35 ttgggcaacg cgaaggcggc cagcgtcgtg gtgactcagg accccatacc ggcttctctg 6120
 gcccaccac actaaccag gaagtacgg agctctgaac ccgtggaaac gaacatgacc 6180
 cttgcctgcc tgcctccctg ggtgggtcaa gggtaatgaa gtgggtgtgca ggaatggcc 6240
 atgtaaatga cagactctg ctgatgggga ccgttccttc catcattatt catcttcacc 6300
 cccaaggact gaatgattcc agcaacttct tcgggtgtga caagccatga caaaactcag 6360
 40 tacaacacc actcttttac tagggccaca gacacggsc cacaccctg atatatgaag 6420
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggctgaac 6480
 agtctgttcc tctagactag tagacctgg caggcactcc cccagattct agggcctggt 6540
 tgctgcttcc cgaggcgccc atctgcccgt gagactcagc ctgggggtgccc acactgaggc 6600
 cagccctgtc tccacaccct ccgcctccag gcctcagctt ctccagcagc ttcttaaac 6660
 ctgggtgggc cgtgttccag cgtactgtc tcacctgtcc cactgtgtct tgtctcagcg 6720
 45 acgtagctcg cacggttctt cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780
 gcgttgaagg gaggagattc tgcgcctccc agactggctc cctctgagct gaacctggct 6840
 cgtggcccc gatgcagggt cctggcgtcc ggctgcacgc tgacctccat ttccaggcgc 6900
 tccccgtctc ctgtcatctg ccggggcctg ccgggtgtgtt cttctgtttc tgtgtcctt 6960
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggtctc ggggttttta taggcatagg 7020
 acgggggctg ggtgggccag ggcgctcttg ggaatgcaa catttgggtg tgaagttagg 7080
 agtgctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
 cccgccttcc tctgccagc actttcctgc cccctccct ctggaacaca gattggcagt 7200
 ttccacaagc actaagcatc ctcttcccaa aagaccagc attggcacc ctggacattt 7260
 55 gcccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320
 ccgaccccc ctgttttatt ttaatagcta caaagcagg aaatccctgc taaaatgtcc 7380
 tttaacaaac tgggttaaca aacgggtcca tccgcacggg ggacagttcc tcacagtga 7440
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgtg gagtcaaaac 7500
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560
 aggggagtggt ttaggggggt taaggacggg gggggcgcca gctgggggct actgcacgca 7620
 60 ccttttacta aagccagttt cctggttctg atggtattgg ctcagttagg ggagactaac 7680
 cataggggag tggggatggg ggaacccgga ggtgtgcca tctttgccat gcccgagtgt 7740
 cctgggcagg ataatgtct agagatgccc acgtcctgat tccccaaac ctgtggacag 7800
 aaccgcccc gccccagggc ctttgacagt gtgatctccg ttaggacctt gaggtctggg 7860
 atccttcggg actacctgca ggcccgaag ggttctggga agaggcggc 7920
 65 aggagggtca gaggggggca gcctcaggac gatggaggca gtcagtctga ggctgaaaag 7980

5 ggagggaggg cctcagcccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040
gggaccctcc acggagcctg cagcaggaag gcacggctgg cccttagccc accagggccc 8100
atcgtggacc tccggcctcc gtgccatagg agggcactcg cgctgccctt ctacgatgaa 8160
gtgtgtgggg atttgcagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
5 aaaacaaagg ttacagaaa catccaagga cagggctgaa gtgcctccgg gcaagggcag 8280
ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340
gttatgctct tgttgcccag gctggagtgc agcggcatga tcttggtca ctgcaacctc 8400
cgtctcctgg gtccaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
gtgcaccacc acaccgggt aattttgtat ttttagtaga gatgggcttt caccatgttg 8520
10 gtcaagctga tctcaaatc ctgacctcag gtgatccgcc cactcagcc tcccaaatg 8580
ctgggattac aggcattgag cactgcacct ggctatttta accattttta aacttcctg 8640
ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700
gttaccctcc tttgatattt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760
catattcaca gtttctgtga ccacctgtta tcccatggga cccactgcag gggcagctgg 8820
15 gaggtgcag gcttcaggtc ccagtgggtg tgcctctgc cagtagaac ctgatgtaga 8880
atcagggcgc aagtgaggac actgtcctga atctcaatgt ctcatgtgt gctgaaacat 8940
gtagaaatta aagtcacatc ctccactct actgggattg agccccttc ctatcccccc 9000
ccaggggcag aggagttcct ctcaactcgt tggaggaagg aatgatactt tttattttt 9060
cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgtttgt tttgagaggc 9120
20 ggtttcactc ttgttgctca ggctggagg agtgcaatgg cgcgatcttg gcttactgca 9180
gcctctgcct ccaggttca agtgattctc ctgcttccgc ctccatttg gctgggatta 9240
caggcacccg ccacatgcc cagctaattt tttgtatttt tagtagagac gggggtgggt 9300
gggggttacc atgttgcca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360
tctgcctcct aaagtgtcg gattacaggt gtgagccacc atgcccagct cagaatttac 9420
25 tctgtttaga aacatctggg tctgaggtag gaagtcacc cactcaagt gttgtggtg 9480
tttaagccaa tgatagaatt tttttattgt ttttagaaca ctcttgatgt ttactactgt 9540
gatgactaag acatcatcag cttttcaaa acacactaac tgacccata atactgggtg 9600
gtcttctggg tatcagaat ctctattgaa tgcgggagg cgtttctcgc ccatgcacat 9660
gggtgtaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaata 9720
30 tgtgttttct atgttggtt ctctgcagg aaccagtgtg agctacaact taacttttgt 9780
tggaaacaaat ttccaaacc gcccttttgc cctagtggca gagacaattc acaaacacag 9840
ccctttaaaa aggtctagg atcactaagg ggatttctag aagagcgacc tgtaatccta 9900
agtatttaca agacgaggc aacctccagc gagcgtgaca gccagggag ggtgcgaggc 9960
35 ctgttcaaat gctagctcca taaataaagg aatttccctc ggagtttct gaaagttaga 10020
aaggttacat ttaaggttgc gtttgttagc atttcagtgt ttgcccagct cagctacagc 10080
atccctgcaa ggctcggga gaccagaag tttctcgcct ccttagatcc aaacttgagc 10140
aaccggaggt ctggattcct gggaagtcc cagctgtcct gcggttgtgc cggggcccca 10200
ggctctggag ggaccagtgg ccgtgtggct tctactgctg ggctggaagt cgggctcct 10260
40 agctctgcag tccgaggctt ggagccaggt gcctggacc cagaggtgcc ctccaccctg 10320
tgcggggcgg atgtgaccag atgttgccct catctgccag acagagtgc gggggccagg 10380
gtcaaggccg ttgtgctggt tgtgagcgc ccggtgcgc gccagcagga gcgcctggct 10440
ccatttccca ccttttctgc acgggaccgc cccgggtgggt gattaacaga tttgggggtg 10500
tttgctcatg gtggggacc ctgcgcctc gagaacctgc aaagagaaat gacgggctg 10560
tgtcaaggag ccaagtgcg ggggaagtgt tgcagggagg cactccggga ggtcccgcgt 10620
45 gcccgtccag ggagcaatgc gtccctgggt tgcctccag ccgcgtctac gcgcctccgt 10680
cctccccctc acgtccggca ttcgtggtgc ccggagcccg acgccccgcg tccggacctg 10740
gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800
cccagggcct ccacatcatg gccctccctc cgggttacc cagagcctag gccgattcga 10860
50 cctctctccg ctggggccct cgctggcgct cctgcacctt gggagcgca gcggcgcgcg 10920
ggcggggaag cgcggcccg acccccgggt ccgcccggag cagctgcgct gtccgggcca 10980
ggccgggctc ccagtggatt cgcgggcaca gacgcccagg accgcgctcc ccacgtggcg 11040
gagggactgg ggaccgggc acccgtcctg ccccttacc ttccagctcc gcctcctccg 11100
cgcggaaccc gccccgtccc gacccctccc ggggtccccg cccagcccc tccgggacct 11160
cccagccct ccccttctct tcccgccccc cgcctctccc tccggcgcg agtttcaggc 11220
55 agcgtgcgt cctgctgcgc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4
<211> 104
<212> DNA
60 <213> Homo sapiens

<400> 4
gtgggcctcc ccggggtcgg cgtccggctg ggggttaggg cggccggggg gaaccagcga 60
catgcggaga gcagcgagg cgactcagg cgcttcccc gcag 104

7 / 18

<210> 5
<211> 8616
<212> DNA
<213> Homo sapiens

5

<400> 5

gtgaggagggt ggtggccgtc gaggggcccag gccccagagc tgaatgcagt aggggctcag 60
aaaagggggc aggcagagcc ctggctcctcc tgtctccatc gtcacgtggg cacacgtggc 120
ttttcgctca ggaagtcgag tggacacggg gatctctgcc tctgtctctcc ctectgtcca 180
gtttgcataa acttacgagg ttcaccttca cgttttgatg gacacgcggg ttccaggcgc 240
cgaggccaga gcagtgaaca gaggagggtg ggcgcggcag tggagccggg ttgccggcaa 300
tgggagagaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360
tcctcttcgc aatttcaagg gtgggaatga gaggtgggga cgagaacccc ctcttcctgg 420
gggtgggagg taagggtttt gcagggtgcac gtggtcagcc aatattgcagg ttgtgttta 480
15 agatttaatt gtgtgttgac ggccagggtgc ggtggctcac gccggtaatc ccagcacttt 540
gggaagctga ggcagggtga tcacctgagg tcaggagtgt gagaccagcc tgaccaacat 600
gggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcatgggtg tgtgtgcctg 660
taatccccag tacttgggag gctgaggcag gagaatcact tgaacccagg aggcggaggc 720
tgcagtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactctgt 780
20 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840
actgttctcc agcacagatc ctggctcccat cttaggtat gaagagggcc acatgggagc 900
agaggacagc agatggctcc acctgctgag gaaggagacag tgtttgtggg tgttcagggg 960
atggtgtctg tgggcccctg cgtgtcccca ccctgttttt ctggatttga tgttgaggaa 1020
ctccgctccc agcccccttt tggctcccag tgcctccagg ccctaccgtg gcagctagaa 1080
25 gaagtccgta ttccaccccc tccccacaaa ctccaagac atgtaagact tccggccatg 1140
cagacaagga ggggtacctt ctgggggctc ttttttttct ttttttcttt ttatggtggc 1200
aaaagtcata taacatgaga ttggcactcc taacaccgtt ttctgtgtac agtgcagaat 1260
tgctaactcg gcggtgttta cagcagggtg cttgaaatgc tgcgtcttgc gtgactggaa 1320
gtccctaccc atcgaaacggc agctgcctca cactgctgct ggcctcagggt gaccacgccc 1380
30 agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc ctctgttgag 1440
gagagtttga gttctctgat caggactctg cctgtcattg ctggtctctg acttcagatg 1500
aggtcaccaat ctgcccctgg cttatgcagg gagtgaggcg tgggtccccg gtgtccctgt 1560
cacgtgcagg gtgagtgagg cgttgcctcc aggtgtccct gtcacgtgta gggtagtgta 1620
ggcgcgcccc ccgggtgttc ctgtcccgtg cagcgtgatt gaggtgtggc ccccggtgtg 1680
35 cctctgtcac tgtagggtga gtgagggccc atccccggg tgcctgtgca cgtgtagggt 1740
gagtgaggcg tgggtccccg gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800
gggtgtccct gtcacgtgca gggtagtgta ggcgcggtcc ccgggtgtcc ctctcagggt 1860
tagggtgagt gaggcgcgcc cccagggtgt ccctgtcacg ttaggggtga gtgaggcacc 1920
gtcccctggg tcccctccca ggtatagggt gagttaggca ctgtccccgg gtgtccctgt 1980
40 cacgtgcagg gtgagtgagg cgcggccccc ggggtgtccct ctccagtgca gggtagtgta 2040
ggcgctgtcc ctgggtgtcc ctgtctctgt tagggtgagt gagggtctgt ccccgaggtg 2100
ccttggcggt tgtcacttg agcttgcctc tgaatgtttg ctctttctat agccacagct 2160
gcgcgggttg cccattgcct gggtagatgg tgcaggcgca gtgctgggtcc ccaagcctat 2220
ctttctctgat gctcggtctt tcttggtcac ctctccgttc cattttgcta cggggacacg 2280
45 ggactgcagg ctctcgcttc ccgctgcca ggcactgcag ccacagcttc aggtccgctt 2340
gcctctgttg ggcctggctt gctcaccacg tgcccggcac atgcatgctg ccaatactcc 2400
tctcccagct tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccacgtg 2460
ttgtcggaga catcccagaa agggttctct gtgccctgaa ggaagcaag tcacccacagc 2520
ccccctactt gtccgttttt ctcccagct gccctctgct tgggccccct tgggtgggtg 2580
50 gcaacgcttg tcaccttatt ctgggcacct gccgtcatt gcttaggctg ggctctgcct 2640
ccagtcgccc cctcacatgg attgacgtcc agccacaggt tggagtgtct ctgtctgtct 2700
cctgtctctga gacccacgtg gagggcgggt gtctccgcca gccttcgtca gacttccctc 2760
ttgggtctta gttttgaatt tcaactgatt acctctgacg ttctctatct tccattgtat 2820
55 gttttttctt gggtttattt ttcatctctt ttctagcttc ttagtttagt catgccttct 2880
cctctaagtg ctgccttacc tgcacctgt gttttgatgt gaagtaatct caacatcagc 2940
cactttcaag tgtttctaaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000
ctgtgatttt tttctttgtg cacgctgtgt tttgacgtga aatcattttg atatcagtga 3060
cttttaagta ttctttagct tattctgtga tttctttgag cagttagtta tttgaacact 3120
gtttatgttc aagatatgta gagtatcaag atacgtagag tattttaagt tatcatttta 3180
60 ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240
ggagccttgc tttgtgatct agtgtgtgca tgggtttccag aactgtccat tgtaaatttg 3300
acatcctgtc aatagtgggc atgcatgttc actatatcca gcttattaag gtccagtgca 3360
aagcttctgt ctccctctag atgcatgaaa ttccaagaag gaggccatag tccctcacct 3420
gggggatggg tctgttcatt tcttctcggt tggtagcatt tatgtgaggc attgttaggt 3480
65 gcatgcacgt ggtagaattt ttatcttctt gatgagtga tcttttggag acttctatgt 3540

ctctagtaat ctagtaattc tttttttaa ttgctcttag tactgccaca ctgggcttct 3600
tttgattagt attttccctgc tgtgtctggt ttctgccttt aatttatata tatatatata 3660
tttttttttt ttttgagaca gagtcttggg ctgtcgccca gggtgagtg agtgggtgtga 3720
tcacagggtca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctcctg 3780
5 agtagctgga actgcagaca cgcaccgcta cacctggcta atttttaaat tttttctgga 3840
gacaggggtct tgctgtgttg cccaggctgg tctcaaaactc ttggactcaa gggatccatc 3900
tacctcggtc tcccaaagtg ctgaattaca ggcatgagcc accatgtctg gcctaaatttt 3960
caacactttt atattcttat agtgtgggta tgtcctgtta acagcatgta ggtgaatttc 4020
10 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca tttttttgtc 4080
actagagacc cgctgggtgc actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140
ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtggt tgttgatcct 4200
ctcgttgccct cctgggtcact gggcatttgc ttttatttct ctttgcttag tgttaccctc 4260
tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccct 4320
ggctggagtg taatggcaca atctcggctc actgcaacct ctgcctctc ggttcaagca 4380
15 gttctcattc ctcaacctca tgagttagct ggattacagg cgcccaccac cagccttgcc 4440
taatttttgt attttttaga gagataggct ttcaccatgt tggccaggct ggtctcaaac 4500
tcctgacctc aagtgatctg cccgccttgg cctccacacg tgcctgggatt acaggtgcaa 4560
gccaccgtgc ccggcatacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620
20 tcttgagcaa taagaccttt agtgtatttt agctctggcc acccccagc ctgtgtgctg 4680
ttttccctgc tgacttagtt ctatctcagg catcttgaca cccccacaag ctaagcatta 4740
ttaatattgt tttcgtgtt gagtgtttct gtatgtttgc ccccgccctg cttttctctc 4800
tttgttcccc gtctgtcttc tgtctcaggc ccgctgctg ggttccccct ctttgtcctt 4860
tgcgtgtgtc ttctgtcttg ttattgtctg taaaccccag ctttacctgt gctggcctcc 4920
atggcatcta gcgacgtccg gggacctctg ctatgatgac acagatgaag atgtggagac 4980
25 tcacaggagag ggcggtcatc ttggcccgtg agtgtctgga gcaccacgtg gccagcggtc 5040
cttagccagt gagtgcagac aacgtccgct cggcctgggt tcagcctgga aaaccccagg 5100
catgtcgggg tctggtggct ccgctgggtg gagtttgaaa tcgcgcaaac ctgctgggtg 5160
gcgcccagtc tgacgggtgct gcctggcggt ggagtgctg ctctctccct tctgcttggg 5220
aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280
30 atgtggataa ttttaaaatt tctaggctgg gcgctgggtg tcacgcctgt aatcccagca 5340
ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400
atgatgaaac cccatctgta ctaaaaacac aaaaattagc tgggcgtggt ggcgggtg 5460
tgtaatccca gctactcggg aggctgaggc aggagaattg cttgaacctg ggagttgga 5520
35 gttgcagtg gcccagattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580
ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aattctagta gccacattaa aaaagtaaaa 5640
aagaaagggt gaaattaatg taataataga ttttactgaa gccacagcatg tccacacctc 5700
atcattttag ggtgttattg gtgggagcat cactcacagg acatttgaca ttttttgagc 5760
tttgtctgcg ggatcccggt tgtaggctcc gtgcgtggcc atctcggcct ggacctgctg 5820
40 ggcttcccat ggccatggct gttgtaccag atggtgcagg tccgggatga ggtcgccagg 5880
ccctcagtg gctggatgtg cagtgtccgg atggtgcagg tctgggatga ggtcgccagg 5940
ccctgctgtg agctggatgt tgggtgtctg gatggtgcag gtcagggggt aggtctccag 6000
gcccctgggt agctggaggt atggagtccg gatgatgcag gtcagggggt aggtcgccag 6060
gcccctgctg gagctggatg tgtggtgtct ggatggtgca ggtcagggggt gaggtctcca 6120
ggccctcggt aagctggagg tatggagtcc ggatgatgca ggtccgggggt gaggtcgcca 6180
45 ggccctcggt tgagctggat gtgtggtgtc tggatggtg aggtctgggg tgaggtcacc 6240
aggccctcgg gtgagctggg tgtgctgggt ctggatgggt caggctctgga gtgaggtcgc 6300
cagacgggtc cagaccatgc ggtgagctgg atatcggtg tccggatggt gcaggtctg 6360
ggtgaggttg ccaggccctg ctgtgagttg gatgtgggtg gtcaggatgc tgcaggtccg 6420
50 gctgtgagtc accaggccct gctgtgagct ggatgtgtgg tgtctggatg gtgcaggtct 6480
ggggtgaagg tcgccaggcc cctgcttgtg agctggatgt gtggtgtctg gatggtgcag 6540
gtctggagtg aggtcgccag gccctcgggt agctggatgt gcaggtctca gatggtgcag 6600
gtccgggggt aggtcgccag accctcgggt gagctggatg tgcgggtgtc ggatggtgca 6660
ggctcggagt gaggtcgcca ggccctcgggt gagctggatg tatggagtcc ggatggtgcc 6720
55 ggctcgggggt gaggtcgcca gaccctcgtg tgagctggat gtgcgggtgtc tggatggtac 6780
aggtctggag tgaggtcgcc agaccctgct gtgagctgga tatgcgggtg ccggatggtg 6840
caggtcagggt gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900
caggtccggg gtgaggtcgc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960
gcaggtctgg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020
gcaggtccgg ggtgaggtcg ccaggccctg ctgtgagctg gatgtgcggc gtcgtgatgg 7080
60 tgcaggtctg ggtgtggtg gccaggccct cgggtgagct gaggtatgga gtcaggatga 7140
tgcaggtccg ggtgaggtt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7200
gtgcagtcct ggtgaggtc gccaggccct gctgtgagct ggatgtgctg tatccggatg 7260
gtgcaggtct ggtgaggtt caccaggccc tgcgggtgag tgggtgtgctg gtgtccgggt 7320
gctgcaggtc cgggtgaggt tcgccaggcc ctccgtgagc tggatgtgctg gtgtcccggt 7380
65 gtccggatgg tgcaggtcca ggtgaggtc gctaggccct tgggtgggctg gatgtgctg 7440

5
10
15
20
25
30
35
40
45
50
55
60

gtccggatgg tgcaggctctg gggtagagtc gccaggccctt tggtagagctg gatgtgcggt 7500
gtctgcattgg tgcaggctctg gggtagagtc gccaggccctt tggtagagctg gatgtgtggt 7560
gtccggatgg tgcaggctccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgcgg 7620
tgtctggatg gtgcaggctcc ggggtgaggt agccaaggcc ttcggtgagc tggatgtggg 7680
gtgtccggat ggtgcaggtc cggggtgagg tcgccaggcc ctgagggttag ctggatatgc 7740
gggtgtccga tgggtgcagg tccggggtga ggtcgccagg ccctgctgtg agctggatgt 7800
cgtgtatccg gatggtgcag gtccggggtg aggtcgccag gccctgcagt gactgtggatg 7860
tgctgtatcc ggatggtgca ggtctggcgt gaggctcgca ggccctgcgg ttagctggat 7920
atgcggtgtc ggatggtgca ggtccggggt gaggctcgca ggccctgcgg ttagctggat 8040
gtgcggtgtc cggatggtgc aggtctgggg tgaggctcgcc aggcctctgt gtgagctgga 8100
tgtgtgttat ccggatggtg cagggtccggg gtgaggtcgc caggccctgc ggtgagctgg 8160
atgtgtgta tccggatggt gcaggctctg cgtgaggtcg ccaggccctg cggtagagctg 8220
gatgtgcagt gtacggatgg tgcaggctcc gggtagagtc gccaggccct ggggtgggct 8280
gtatgtgtgt tgtctggatg gtgcaggctcc ggggtgagtt cggcaggccc tgcggtgagc 8340
tggatgtgtg gtgtctggat gctgcaggtc cggggtgagt tcgccaggcc ctgcggtgagc 8400
tggatatgcg gtgtccccgt gtccgaatgg tgcaggctca gggtagagtc gccaggccct 8460
tgggtgggctg gatgtgccgt gtccggatgg tgcaggctcg gggtagagtc gccaggccct 8520
tggtagagctg gatgtgcggt gtccggatgg tgcaggctcg gggtagagtc accaggccct 8580
cggtagatctg gatgtggcat gtccctctctg ttttaag 8616

<210> 6
<211> 2089
<212> DNA
<213> Homo sapiens

<400> 6
gtactgtatc cccacgccag gcctctgctt ctccaagtc tggaaacacca gcccggcctc 60
agcatgcgcc tgtctccact tgctctgctt tccctggctg tgcagctctg ggtcgggagc 120
caggggcccc gtcacaggcc tgggtcaagt ggattctgtg caaggctctg actgcttga 180
gctcacgttc tcttacttgt aaaaacagga gtttgcgca agtggtctct aggggttgta 240
aagcagaagg gattttaaatt agatggaaac actaccacta gcctccttgc ctttccctgg 300
gatgtgggtc tgattctctc tctctttttt ttttctttt tgagatggag tctcactctg 360
ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
ttaagcgatt caccagcctc agcctcctaa gttagctggga ttacaggcac ctgccaccac 480
gcctggctaa tttttgtact tttaggagag acgggggttc accatgttgg ccaggctggt 540
ctcgaaactc tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600
ggctaagcca ccgtgcccag cccccgattc tcttttaatt catgctgttc tgatgaatc 660
ttcaatctat tggatttagg tcatgagagg ataaaatccc acccacttgg cgactcactg 720
cagggaagcac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780
taggtgggctg catttgaatg gctgtgagat tttgtctgca atgttcggct gatgagagtg 840
tgagattgtg acagattcaa gctggatttg catcagttag ggacgggagc gctgggtctg 900
gagatgccag cctggctgag cccaggccat ggtattagct tctcctgtgc ccgccaggc 960
tgactgtgga gggctttagt cagaagatca gggcttcccc agctccctg cactctgag 1020
tccctggggg gccttgtgac accccatgcc ccaaatcagg atgtctgcag agggagctgg 1080
cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgacg tgggtgctgg 1140
gccatttccct tgcactctgg ggagggtcag ggccttccct gtgggaacaa gtttaatacac 1200
aatgcacctt acttagactt tacacgtatt taatggtgtg cgacccaaca tggctatttg 1260
accagtattt tggaaagaat ttaattgggg tgaccggaag gacagacag acgtggtggt 1320
ccccaaagat ctccctgtca ctactgggac tgtgttctg cctggggggc cttggagggc 1380
cctcctccct ggacagggtg ccgtgccttt tctactctgc tgggctgctg gctgctgctc 1440
agggcaccag ctccggagca cccggggccc cagtgtccac ggagtgcag gctgtcagcc 1500
acagatgcc aggtccagg gtggccgctc cagccccgt gccccatgg tgggttttg 1560
gggaaaaggc caagggcaga ggtgtcagga gactggtggg ctcatgagag ctgattctgc 1620
tccttggctg agctgcccct agcagcctct cccgcctct ccatctgaag ggatgtggct 1680
ctttctacct gggggtcctg cctggggcca gccttgggct acccagtggt ctgtaccaga 1740
gggacaggca tcctgtgtgg aggggcatgg gttcacgtgg cccagatgc agcctgggac 1800
caggctccct ggtgctgatg gtgggacagt caccctgggg gttgaccgac ggactgggac 1860
tcccagggt tgactatagg accaggtgtc cagggtgcct gcaagtagag gggctctcag 1920
agggctctgt ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccc 1980
tgggtgccct gagccctcac tgagtcgggt ggggctgtg gcttcccgtg agcttcccc 2040
tagtctgttg tctggctgag caagcctcct gaggggctct ctattgcag 2089

10 / 18

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5

<400> 7
gtggctgtgc tttggtttaa ctctcttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccggtgcc tccagaaaag 180
cagcgtgggg gtgtaggggg agctcctggg gcaggagacag gctctgagga ccacaagaag 240
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tccctgctgtg 300
gtgcgcagcc tccgtgcgct tccgcttacg gggcccgggg accaggccac gactgccagg 360
agccaccggg gctctgagga tccctggacct tgccccacgg ctccctgcacc ccaccctgt 420
ggctgcggtg gctgcggtga ccccgctatc tgaggagagt gtggggtgag gtggacagag 480
15 gtgtggcatg aggatcccg gtgcaacaca catgcggcca ggaaccggt tcaaacaggg 540
tctgaggaaag ctgggagggg ttctaggtcc cgggtctggg tggctgggga cactggggag 600
gggtgcttc tccctgggt ccctatggtg ggggtggcac ttggccggat ccactttcct 660
gactgtctcc catgtgtcc cggccag 687

20

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

25

<400> 8
gtgggtgccg gggacccccg tgagcagccc tgctggacct tgggagtggc tgcctgattg 60
gcacctcatg ttgggtggag gaggtactcc tgggtgggccc gcaggagagt cagggtgaccc 120
tgctactgtt gaggacacac ctggcaccta ggggtggaggc ctccagcctt tccctgagca 180
catggggccg actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtcccactg 240
30 gattccagt tccgtcagag aaggaaaccg aacggctcag ccaccaggcc ccggtgcctt 300
gcaccccagt cctgagccag gggctcctg tccctgaggct cagagagggg acacagcccc 360
ccctgccctt ggggtctgga gtgggtgggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag tttctgctg gccactgtca gtctcctcgc 480
ctccactcac acag 494

35

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

40

<400> 9
gtaagggtca cgtgtgatag tctgtgccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtgggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggtgcat gtatctgtgg cgtgcataat 180
45 tgtgtgtgtg gtgtgtgtgg cactgtgtgt tccatgggtg gtgtgcctgt ggtgtgcatg 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgccta 300
tttgtgtgtg gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360
ggcccccttg ccttactcct tccctcctca ggcattgtcc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggctcctc cttctagcat gggtgcccc 480
50 gtcctgtcac agggctgggc cttggagact gtaagccagg tttagagga gaggtagggat 540
gctggtggtg ccttctctga cccctggcac cccaggacc ccagtctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctgcgtcccc gggacacact cctcccagag 660
cggccggggg ccttggggct cggcaggggt gaaaggggccc ctgggcttgg gttcccaccc 720
agtgtgcatg agcacgctgg aggggtaaag cctcaaagtc gtgccaggcc ggggtgcaga 780
55 ggtgaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgacttctt gagct 865

60

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

<400> 10
tgtgggattg gttttcatgt gtgggatagg tggggatctg tgggattggg ttttatgagt 60
ggggtaacac agagttcaag gcgagctttc ttctgtagt gggctctgag gtgctccaac 120

agctttattg aggagacccat atcttccctt gaactatggt cgggtttata gtaagtcagg 180
 ggtgtggagg cctccccctg gctccccctt ctgtttcttc cactctgggg tcgtgtgggtg 240
 cctgtctgtg tgtgtggccg gtgggcaggg cttccaggcc tccttgtgtt cattggcctg 300
 gatgtggccc tggctacgct ccgtcccttg aattccccct cgagttggag gctttctttc 360
 5 tttctttttt tttttctttt tttttttttt tgataacaga gtctcgctct tttttgccc 420
 ggctggagtg gtttggcgtg atcttggctc actgcaacct gtgcttccct agttcaagca 480
 attctcttgc ctcagcctcc caagtagctg gaattatagg cggccaccac catgctgact 540
 aattttttgt atttttagtag agacgagggt tctccatgtt ggccaggctg gtctcgaact 600
 cctgacctca ggtgatcctc ccacctcggc ctcccaaagt gctgggatga cagggtgtgaa 660
 10 ccgcccgccc cggccgagac tcgcttccct cagcttccct gagatctgca gcgatagctg 720
 cctgcagcct tgggtgctgac aacctccgtt ttccttctcc aggtctcgct aggggtcttt 780
 cactttcatg actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgtttcct 840
 gcgtaattgg tgtctgctgt ttatcgatgg cctccttcca tttcctttag gctttgttta 900
 ttgtgttttt tccggctcct tgaagggaaa gtttcgatta tggatgtttg aactttcttt 960
 15 tctaaacaag catctgaagt tgccgttttc cctctaaagc agggatcccg agggccctgg 1020
 ctgtggagtg gcaccggtct ggggcctgtt aggaaccggg cgacacagcg gaggctaggt 1080
 ggggtgtggg gagccagcgt tcccgccctga gcccgcctcc tctcagatca gcagtggcat 1140
 gcggtgctca gaggcgacac caccctactg agaactgtgc gtgagagggg tctagattct 1200
 gtgtccttta tgggaatcta atgctgatg atctgaggtg gaaccgtttg ctcccaaac 1260
 20 catccccctc cccactgctg tctgtgtgaa aaactcgttt ccacgaaacc agtccctggg 1320
 accacaatgg ttggggaccc tgtgctaaag acctgcttca gcagcctctc gtcagtgttg 1380
 atatatggc ttttctgtgt tgagtcacga ataattacgg atttctgtga tgctttccgc 1440
 cgacctcaga cccatgggct atttgtgggc gtgttgccct ctccctgggtt gggaagggtg 1500
 caggccccat gtaccttctt gttactgcct tccagggttg ttctcagggt tgaatcgta 1560
 25 tcgatgtggt tttagcccac ggccctgccc ccagctcctg ggggctgggg aacatgctga 1620
 agcacagagt caccgtgcgc gtcttttcat gcctcacaag ctcgaggcct cctgtgtccg 1680
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtcttggggg agtctgcaga ataggaggtg ggggtgccgg 1800
 tctctctccc gcgtcttcag actcttctcc tgcctgtgct gtggctgcac ctgcatccct 1860
 30 gcaatccctc cagcactggg ctggagaggg ccgggagctc gactgcccact tgtgccacgt 1920
 gactgtggat ggcagtcggg cacgggggtc tgatgtgtgg tgactgtgga tggcggttg 1980
 tcacaggggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgact 2040
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga 2100
 tgtgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggctcgt 2160
 35 ggggtctgact tgggtgactgt ggtatggcagt cgtggggtct gatgtgtggt gactgtggat 2220
 ggcggctcgt ggggtctgat tgggtgactgt ggtatggcagt cgtggggtct gatgtgtggt 2280
 gactgtggat ggcggctcgt ggggtctgat tgtgtgtgact gtggatggcg gtcgtgggt 2340
 ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgga ctgtggatg 2400
 cggtcgtggg gtctgatgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgga 2460
 40 ctgtggatgg tgatcggtca cagggtctcg atgtgtgtg actgtggatg gcggtcgtg 2520
 ggtctgatgt gtgtgactg tggatgggtga tccgtcacag gggctctgat tgtgtgtgact 2580
 gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggttg gtcccgggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtgactgt 2700
 ggatggcggt cgtggggtct gatgtgtggt gactgtggat ggcggtcgtg gggctctgat 2760
 45 tgtgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggtcgtg 2820
 ggggtctgat tgggtgactgt ggtatggcggt cgtggggtct gatgtgtggt gactgtggat 2880
 ggcgggtggt cccgggggtc tgatgtgtgg tgactgtgga tggcggtcgt ggggtctgat 2940
 gtgtgtgact tggatggcag tctgtgggtc tgatgtgtgg tgactgtgga tggcggtcgt 3000
 ggggtctgat gtgtgtgtgac tgtgtgtggc ggtctgtggg tctgatgtgt ggtgactgtg 3060
 50 gatggcggtc gtggggtctg atgtgtgtg actgtggatg gcggtcgtgg ggtctgatgt 3120
 ggtgactgtg gatggcggtc gtggggtctg atgtgtgtg actgtggatg gtgatcggtc 3180
 acaggggtct gatgtgtgtg gactgtggat ggcggtcgtg ggggtctgat tgtgtgtgact 3240
 gtggatggcg gtctgtgggt ctgatgtgtg gactgtggat ggcggtcgtg ggggtctgat 3300
 tgtgtgtgact gtggatggcg gtctgtgggt ctgatgtgtg gtgactgtgg atggcggtc 3360
 55 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtggggtctg atgtgtgtg 3420
 actgtggatg cgggtcgtgg ggtctgatgt gtgtgtgact tggatggcg tctgtgggtc 3480
 tgatgtgtgg tgactgtgga tggcggtcgt ggggtctgat gtgtgtgact tggatgtgga 3540
 tcggtcacag ggggtctgat tgtgtgtggt gcaggtggag tccaggtgt gtctgtagct 3600
 actttgcgtc ctgcggcccc cggccccctt tcccaaacaa gaagcttccc aggcgtctc 3660
 60 tgggtctcat cccgccatcg ggtctggcgg cagggtccaca cgtcctgac ggaagaaaca 3720
 agtgcaccag tctgtccggg gcaggccaca tttgtggctc atgccctctc ctctgccggc 3780
 ag 3782

12 / 18

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

5

<400> 11
gtctggggcac tgccttgcag ggttggggcac ggactcccag cagtgggtcc tcccctgggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaatgagc 120
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180
10 tgcgacagct gctgcattca ggcacctgct cactgttgac tgcgcgccct ctctccagtt 240
ccgcagtgcc ttgtttcatg atttgctaaa tgtcttctct gccagttttg atcttgaggc 300
caaaggaaaag gtgtcccccct cctttagggag ggcaggccat gtttgagccg tgtcctgccc 360
agctggcccc tcagtgtctg gtctgaggcc aaaggaaaac tgtccccctt cttaggaggga 420
cgggctgtgt ttgagccacg ccccgctgag cgggcctctc agtgcctggg ctgtccacgt 480
15 ggccctgtgg ccttttgcag atgtggtctg tccacgtggc cctgtggctc ttgacagatg 540
cctgttagca ctgtctggc tctaggggac agtcgtgtcc accgcagtag gctcagagac 600
ctctggcgca atttcttgg ctcccagggt gggggtggag gtggcctggg ctgctgggac 660
ccagaccctg tgcgcggcag ctgggcagca actcctggat cacatatgcc atccgggcca 720
cgggtgggtg tgtgggtgtg agcccagctg gacccacagg tggcccagag gagacgttct 780
20 gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccg gccagccac 840
gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaggga 900
cggaggggtct tggccacgtg gtctgctg tctcagcacc caccggctca ctcccatgtg 960
tctcccgtct gctttcgcag 980

25

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

30

<400> 12
gtgagtcagg tggccagggt ccattgccct gcgggtggct gggcgggctg gcagggtctc 60
tgctcacctc tctcctgccc cttcccact gnccttctgc ccggggccac cagagtctcc 120
ttttctggcc ccgccccct ccggtcctg ggtgcagggc tcccagaggc ccggaacat 180
ggctcggctt gcggcagccg gaggcgagca ggtgccacac gaggcctgga aatggcaagc 240
35 ggggtgtgga gttgctcctg cgtggaggac gaggggcggg ggggtgtgtc gggtcaggtg 300
tgcgcccagc gtttgagcct gcagctgtc agctccaaat tactactgac gctggacacc 360
cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420
tctgttccct gtcgtgtgac ccccgcgagg gcgcgggctc tctctctgt gactagattt 480
cccatctgga aagtgcgggg ttgaccgtgt agtttgcctc tctcgggggg cctgtggtgg 540
40 ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600
cgggtgttaga gccacagtgc ctggtgccac atcactgctc ctggatttta agtaaaacca 660
cacacctccc ggcaggcatc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacact caaggctatc agcaaggta tccgcagta ggtggaacgt 780
ggaggcctct ctctgggacg gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840
45 tttattttaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaaagta gtacacacgt tctggaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020
taagcgcccc ccaggccccc agaattcgtc gacaaagta cctccccaga gaagccacca 1080
cgggcctcct tcgtggtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
50 gtggcaggcg tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200
ggtgactgtg tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtcccatcg 1260
tggctccagt tggcctctga ataaaaacgt cttcaaaacc tgttgcccca aaaactaaga 1320
acagagagag tttcccatcc catgtgtca caggggcgta tctgcttgcg ttgactcgtc 1380
gggtgtggcg gactcctaga gttggtgctg gtgcttctgt gcaaaaagtg cagtcctctt 1440
55 gcccatact gtgatattc caccagcaag gaaagcctct tttcttttct tctttttttt 1500
ttttttgaga cggaacgtca ctgtgtgtct cctgggcttg agtgacagtg cgcatctca 1560
actcactgca acctccgct cccgggttcc agcatttctc ctgcctcagc ctcccagca 1620
gctgagatta caggcaccca cccctgctgc ctggctaatt tttgtatttt tagtagagag 1680
gggtttttgc catgttggtc aggtgtgtct cgaactcctg acctcaggtg atccaccac 1740
60 ctgcgctccc caaagtgtgt ggattacagg tgtgagccat cagccccagc cggaagcct 1800
cttttaagg tgaccaccta tagcgcttcc cgaaaataac aggtcttgtt ttgtagtag 1860
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctctgggga tggctgaggg 1920
tcgctgtgga gccatgcctt ctgtgtgcac ctttaggttc cagggggcta tctgtctctc 1980
actgtttgtc tgaaaacgca cccttggcat ccttgttttg agagtttctg cttctcgttg 2040
65 gtcagtctga aactaggggc aaggtgtgat ccgttggcgc gcagcggcta catgtagggt 2100

13 / 18

	catgagtctt	tcaccgtgga	caaattcctt	gaaaaaaaa	aaaggagtcc	ggtaagcat	2160
	tcattcccgg	tcaagtgtct	ggttctgtga	ataaactcta	agattttaaga	aaccttaaty	2220
	aaagaaaacc	ttgatgattc	agagcaagga	tgtggtcaca	cctgtggctg	gatctgtttc	2280
	agccgcccc	gtgcatggtg	agagtgggga	gcagggtattg	tttgttcaga	ggtctcatct	2340
5	ggtagtcttc	tgaggtgttt	gccggctgaa	tggtagacgt	gtcgtttgtg	tgtatgaggt	2400
	tctgtgtctg	tgtgtggctc	ggtttgagtg	tacgcattgc	cagcacatgc	cctgcccgtc	2460
	tctcacctgt	gtcttcccgc	cccag				2485
	<210> 13						
10	<211> 1984						
	<212> DNA						
	<213> Homo sapiens						
	<400> 13						
15	gtgaggcctc	ctcttcccc	ggggggcttg	ggggggggt	gatttgcttt	tgatgcattc	60
	agtgttaata	ttcctgggtg	tctggagacc	atgactgtct	tgtcttgagg	aaccagacaa	120
	ggttgcagcc	ccttcttggt	atgaagccgc	acgggagggg	ttgcacagcc	tgaggactgc	180
	gggtccacg	caggctctgt	ccagcggcca	tgtccagagg	cctcagggct	cagcagggcg	240
	gagggccgct	gccctgcatg	atgagcatgt	gaattcaaca	ccgaggaagc	acaccagctt	300
20	ctgtcacgtc	acccaggttc	cgttaggggc	cttggggaga	tggggctggg	gcagcctgag	360
	gccccacatc	tcccagcagg	ccctcgacag	gtggcctgga	ctggggcgct	cttcagcccc	420
	ttgccccatc	cacttgcatg	gggtctacac	ccaaggacgc	acacaccta	atatcggtgc	480
	aacctaagt	ggttcaactc	agctggcttt	tattgacagc	agttactttt	tttttttaa	540
	tactttaagt	tctagggtag	atgtgcacga	cgtgcagggt	agttacatat	gtatacatgt	600
25	gccatgttgg	tgtgtgcac	ccattaactc	atcatttaca	ttaggtatat	ctcctaagtc	660
	tatccccccc	cactcccccc	atccccgac	aggccctggg	gtgtgatgtt	ccccaccctg	720
	tgtccaagtg	ttctcattgt	tcagttcccc	cctgtgagtg	agaacatgtg	gtgtttgggt	780
	ttctttccct	gcaatagttt	gctcagagtg	atggtttcca	gcttcgtcca	tgtccctaca	840
	aaggacatga	actcctcctt	ttttatgact	gcatagtatt	ccgtggtgta	tatgtgccac	900
30	attttcttaa	tccagtctat	catcgatgga	catttgggtt	ggttgcaagt	ctttgctact	960
	gtgaatagtg	ccgcaataaa	catacgtgtg	catgtgtctt	tatagcagca	tgatttataa	1020
	tcctttgggt	atatacccag	taatgggatg	gctgggtcaa	atgggtattc	tagttctaga	1080
	tccttgagga	atcaccacac	tgtcttccac	aatgggtgaa	ctagtttaca	ctcccaccaa	1140
	cagtgtataa	gtgttctggt	gctggagagg	atgtggacag	cagttatttt	tttatgaaaa	1200
35	tagtatcact	gaacaagcag	acagtttagt	aaggatgcgt	caggaagcct	gcaggccaca	1260
	cagccatttc	tctcgaaagc	tccgggtttt	tcctgtgcat	cttttgaaac	tctagctcca	1320
	attatagcat	gtacagtggg	tcaaggttct	tcttcattaa	ggttcaagtt	ctagattgaa	1380
	ataagtttat	gtaacagaaa	caaaaatttc	ttgtacacac	aacttgctct	gggatttgga	1440
	ggaaagtgtc	ctcgagctgg	cggcacactg	gtcagccctc	tgggacagga	tacctctggc	1500
40	ccatgggtcat	ggggcgctgg	gcttgggcct	gagggtcaca	cagtgcacca	tgcccagctt	1560
	cctgtggata	ggatctgggt	ctcggatcat	gctgaggacc	acagctgcca	tgctggtaaa	1620
	ggggcaccag	tggctcagag	ggggcgaggt	tcccagcccc	agctttctta	ccgtcttcag	1680
	ttatttttcc	ctaagagtct	gagaagtggg	gccgcgcctg	atggccttcg	ttcgtcttca	1740
	gctggcacag	aattgcacaa	gctgatggta	aacactgagt	acttataaat	aatgaggaat	1800
45	tgcgttagca	gttaactgta	gagagctcgt	ctgttggaaa	gaaatttaag	tttttcattt	1860
	aaccgctttg	gagaatgtta	ctttatttat	ggctgtgtaa	attgtttgac	attcagctcc	1920
	tcgtagacag	atactacgta	aaaagtgtaa	agtttaacct	gctgtgtatt	ttcccttatt	1980
	ttag						1984
	<210> 14						
50	<211> 1871						
	<212> DNA						
	<213> Homo sapiens						
	<400> 14						
55	gtgaggcccc	tgccgtgtgt	ctgtggggac	ctccacagcc	tgtgggcttt	gcagttgagc	60
	cccccgctgc	ctgcccctgg	caccgcagcg	ttgtctctgc	caagtcctct	ctctctgccc	120
	gtgctggatc	cgcaagagca	gaggcgcttg	gccgtgcacc	caggcctggg	ggcgaggggg	180
	caccttcggg	agggagtggg	taccgtgcag	gccctgtgtc	tgagagagcg	caccaggttt	240
60	acacacgtgg	tgagtgcagg	cggtgacctg	gctcctgctg	ctctttggaa	agtcaagagt	300
	ggcggtcctt	ggggccccag	tgagaccccc	aggagctgtg	cacagggcct	gcagggccga	360
	ggcggcagcc	tcctccccag	ggtgcacctg	agcctgcgga	gagcaggagc	tgctgagtga	420
	gctggccccc	agcgttcgct	gcgggtcacgt	tcctgcgtgg	ggttgttttg	gatcggtggg	480
	agaatttggg	tttgcgtgag	gctgctgtct	tgaaccacgg	agatggctag	gagtggtttt	540
65	cagagttagt	ttttgtgaat	caaaactaaa	tcaggcacag	gggacctggc	ctcagcacag	600

5 gggattgtcc aatgtggtcc ccctcaaggg cgccccacag agccggtggg cttgttttaa 660
 agtgcgattt gacgagggac gaaaaacctt gaaagctgta aaggggaacc tcagaaaatg 720
 tggccgccag ggggtggttc agtgctttt ctgggctgtg tttgtgaaaa ccattttgga 780
 cccgccctcc aagtccaccc tccaggtcca ccctccaggg ccgccctggg ctgggggtat 840
 10 gcctggcggt ccttgtgccc cagcccggag cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cgggggggag cagggtccca agcaactgag ggctcaggag tcttgaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtctc tgagggtgct 1020
 ggccagggag gtggctcaga gtgtatgttg ggggtcccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggg gtggccacag ggagctggga atgcaccagg 1140
 15 ggagctgccc agctggccga ggtcccaggg ccaggccaca ggaaggcag ggggacgccc 1200
 ggggccacag cagaggcccg aggaagggaa ggggatgccc aggccagagc agaggctacc 1260
 gggcacaggg gggctccctg agctgggtga gcgaggctca tgactcggcg agggaaacctc 1320
 cttgacgtga agctgacgac tgggtttgcc cagctcacag ccagccagg tcccgcgcct 1380
 gagcaggaac tcagaaacct cccctttgtc taaagcacag cagatgcctt cagggcactc 1440
 20 aggagaaaaa aggcaaaagc gttgagaaa gctttaaag aagggtggat ggtggcaatt 1500
 tcttgtccag atttttagtct gccccggacc acagatgagt ctataacggg attgtggtgt 1560
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccactc 1620
 gagtccctggc tgtcccgggt ccaggccagg ttcttgcatg ctacactacc tgtcctgccc 1680
 gggagacagg gaaagcaccc cgaagctgg agcagggtg ggtccaggct cctcagagct 1740
 25 cctgccaggc ccagcacctc gctccaaatc accactctc tggggtttc caaagcattt 1800
 aacaagggtg tcaggttacc tcttgggtga cggccccgca tcttggggct gacattgccc 1860
 ctctgcctta g 1871

<210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens

30 <400> 15
 gtgagcgac ctggccggaa gtggagcctg tgcccggctg gggcagggtg tgctgcaggg 60
 cegtgtcgct cacctctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggg tgcccctcgt cccatctggg gctgagcaga aatgcactt tctgtgggag 180
 tgaggggtgt cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240
 35 cagacctggg tgcactgagg tgccttcaga aagcagctct gatccgaacc caagacgccc 300
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg ggcctctgtg ggcagtagtc 360
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgccatgag ttcatgatca cgtgtgacct atcaggggac agggccatgg 540
 40 tgtggggggg gtctctacaa aattctgggg tcttgtttcc ccagagccc agagctcaag 600
 gccccgtctc aggtctcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa agtatcaac attccaggca gggcaagggt gctcacacct 720
 ataattccag cactttggga ggccgagggt ggtggatcac ttgaggccag gagtttgagg 780
 ccaacctaac caacatagt aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 45 cctggtggca cagcctgta gtccccgcta tgcgggaggc tgaggcagga gaatcatttg 900
 aaccaggag gcagaggttg cagtgagccc agatcacacc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
 ggacaggtgt ttttttatcc tgccttcga taatatttac tgggtgctgtg ctagaggccg 1080
 gaaactgggg tgcccttctc tgaaggcac accttcatgg gaagagaaat aagtgggtga 1140
 50 tgggtgttaa accagaggtt taaactgggg tctgtcgttt ctgagttaac agtccagatc 1200
 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcagggtg 1260
 tggacaccct cgtgatgggg gagcagcagg tgcagacgcc ctcatgatgg gggagtggca 1320
 ggtgcagaca cccttgtgca tgggtgcccg catgtccctg ttgcagctcc ctccccacaa 1380
 ggatgccggc ctctctgtgt ccccacagtc cctgcttccc tctcacagcc ttacctgggtc 1440
 ctggcctcca ctggctttgt ctgcatgatt tccacatttc ctgggctccc agcactctct 1500
 55 cgctctccc aggcacctct gcagtgtggt ccataccagt cagctgtgaa ctgtccactg 1560
 ctatttttgc tccccatgaa atgtatttt taggacaggc acccctgggt ccagcctctg 1620
 gcacagcatc agtgaatggt attgaaggac aaaggacaga caaacaatc aggaaaatgg 1680
 gttctctcta aacacattgc aaagccacag aggctagtgc aggatgggtg ggcacagggt 1740
 catcagatgt ggggtccaat ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800
 60 gtgtgcttgc agaggtgggt ctaaaagctc agcagtggag gcagtgggtc gccatactca 1860
 ggggtgaact acatcctctg tgtctgaagt atacagcaga ggcttgaaag gcactctgga 1920
 gaagaaaaa ggcaaaatga ttaagaaaag tgaaaaagga aaagtggtaa gatgggaatt 1980
 ttcttgtcca gatttttagtc tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040
 gatggacaga acaatagaac aaaacggaag ccctatctct cagaaaacgtg tgttaattgt 2100
 65 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaact 2160

15 / 18

5 gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220
aggagacaca tgcaaaacaac accagcaaca gaaataaaac aaaagactca aaggggaaggg 2280
aggtgaacgt tccctggttt ggtgttgggg aaggacacac agggaggcgg atgaaaccag 2340
tgaggcaacg ggcattgctt tcaactgcaga gaaactcagc ttgcctgagc cacagtgaag 2400
atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgcctg tgaggtcctg 2460
cacattcattc ctctcacttt gtctctctaa ccacttgaga ggtagaggag gaaaggctcc 2520
aggggagcag cgcctctggg tcacccagct ggcaaaagggc atgcatgatt gcagcctggc 2580
ctcctgctcc ggggcccctg ctctgcccga ggaccccaca caagtacagc ccataggctc 2640
aggggtgagcc ggagcccaag gtctgtgttg ggatggctgt gaaagaagaa atggacgtct 2700
10 gatgcacact tgggaaggtc ctaccagcag cgtcaagaa atgcatgtga aactgacagc 2760
gagacccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820
atgctggctc cttttctggg cttgccaaga gccagcatca ggttgaggca agctggaag 2880
acttttctgg aaagcagctt gtttgatgg aagtcctcac aatgtcctgt gtcttcccag 2940
taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtgtt 3000
15 ccaggcaggg ggaactgcca cagcaagtca cgaacctgcc caaatacagg gctaaaggaga 3060
tattatgcat caaaaaactt gctctgccat taaacatttt tcaaagaatt ttgaagaat 3120
gtttaatggc aaaaaacgtt tatttcaatg tagcagtgtt caaagctgga tgtaaaagaa 3180
cacaccccag gacccctgcg tgaatgtcat gtgtgttcat ctttggacat ggacatacat 3240
20 gggcagtgag tgggtgtgag gccctggagg acatcggttg gatgectcca tccctgcccct 3300
ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360
gctcttccat ccttgagatt caaacacagt gagattcccc acgcccactc cagtgttctc 3420
ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cgcctgggag ccagggtctc 3480
acagtttatt atgtgttttt ggtgtgatta tgtgcagatc tcatcagggc agatgatgag 3540
25 tgcacaaaca cggccgtgag aggtttggat acactcaaca tcaactagcca ggtcctggtg 3600
gagtttggtc atgcagagtc tggatggcat gtacgatttg gagtccatgg agtgagcacc 3660
cagccccctc gggctgcagc gcattgcccc ggcaggacaa ggaagcggga ggaaggcagg 3720
aggctctttg gagcaagctt tgcaggaggg ggctgggtgt ggggagggca cctgtgtctg 3780
acattcccc ctgtgtctca g 3801

30 <210> 16
<211> 880
<212> DNA
<213> Homo sapiens

35 <400> 16
gtgagcagggc tgatggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60
gtgtgtgtgt gtgcgcgcgt gcctgcaagg ctgatggtga ctggctgcac gtaagagtgc 120
acatgtacgc atatacagc gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180
40 cagtgtgtgc acaggtgtgc aagggcacaa gtgtgtgcac atgcgaatgc acacctgaca 240
tgcatgtgtg ttcgtgcaca gtcgtgtggg cattcacagt aggtgcatgc gtgtgggtgt 300
gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttcaccccgc 360
taggtctctca gccacagtgc cactccttac aggatgagac ggggtccccc gccttgggtg 420
gctgaggctc tgaagctgca gccctgaggg cattgtccca tctgggcatc cgcgtccact 480
ccctctcctg tgggcttctg tctccactcc ccctctcctg tgggcattta catccactcc 540
45 actccctctc tctgtgtggc atccgcgtcc actccccctc tctgtgggca tctgcgtcca 600
cctccccctc ctgtgggcat ttgcgtccac tccctctcct ggttccttcc tgtcttggcc 660
gagcctcggg ggcaggcaga tgacacagag tcttgactcg cccagggtgg ttcgcagctg 720
ccgggtgagg gccaggccgg atttacttgg gaagagggat agtttcttct caaaatgttc 780
ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840
50 gagggtttcta ccgtttctca ctctttcttg gcgactctag 880

<210> 17
<211> 3186
<212> DNA
55 <213> Homo sapiens

<400> 17
gtgagccgcc accaaggggt gcaggcccag cctccaggga ccctccgcgc tctgtctacc 60
tctgaccggg ggcttcacct tggaaactct gggtttttag ggcaaggaaat gtcttacgtt 120
60 ttcagtgggt ctgtgctcgt tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180
tctccatctc tgggtagtgg taggagccgg tgggtcccca ggtgtcccca ctgtgctctg 240
gcactggcgg tgggacgtca tggaggccat cccagggcag caggggcatg gggtaaagag 300
atgtttatgg ggagtcttag cagaggaggc tgggaagggt tctgaacagt agatgggaga 360
tcagatgccc ggaggatttg gggctctcagc aaagaggggc gagggtgggtg cagggtgagg 420
65 tgcctggccc cccccccggg aagggtgcagc agagctgtgg ctccccacac agccccggca 480

16 / 18

5 gcacctgtgc tctgggcatg gctgtgtctc tggaaagtcc cctgtcctgg ctgggtcaggg 540
 ggtgccccctg ccaagaatcg acaactttat cacagagggg agggccaatc tgtggaggcc 600
 acagggccag cttctgctcg gagttagggc aggtgggtgg acaagcctcg gggctgtacc 660
 aaagggcagt cgggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720
 ggagctgaat gccaggaggc cgaagccctc gccccatgag ggctgagaag gagtgtgagc 780
 atttgtgtta cccagggccg aggtgctgag aattaccgtg cacacttgat gtgaaatgag 840
 gtcgtcgtct atcgtggaag cccagcaagg gctcacggga gagttttcca ttacaaggte 900
 gtaccatgaa aatgggtttt aaccgagtg cttgcgcctt catgtctctg cagggaggggc 960
 10 agagccacag ctgcatgtta ccgcttttgc accagctcca gaggcttggg accaggctgt 1020
 ctacgttcca ggggtcgtcc gggtcagacc gccctcctct ctgccttctc tctctgcctc 1080
 aaatctttcc tcgtttgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctcttttccg gaaacccttg ggggtgtgctg gatacaggtg ccactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggcctcct tgggcatga 1260
 tgaggtcaga ggagttttcc caggtgaaaa ctccctggga actccaggg ccattgtgacc 1320
 15 tgccacctgc tcctccata ttccagctcag tcttgtctc atttccccc cagggctctc 1380
 agctccgagg agctcccgta gaggccctgg gctcagggca gggcggtgca gtttccccac 1440
 ccattgtggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcca 1500
 tgggcccagg gccgtttcca aacacagagt caggcacgtg gaaggcccag gaatccctt 1560
 20 ccctcgagc agggagtggg gaacggagag ctgggccccg atttcacggc agccaggctg 1620
 cagtgggcga ggctgtgtg gtccacgtgg cgtggggggc ggggtctgat tcaaatccgc 1680
 tggggctcgg ccttctctgg ccgtgctggc cgcgcctcca cagggtctg ggggtgacgc 1740
 cccgacctct agcagggtggc tatttctccc tttggaagag agccctcac ccattgtagg 1800
 tgttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggtctattt 1860
 atttgtttaa aaacattctg ggcctggctt ccgtgtgtgc taaatgggga aaagacatcc 1920
 25 cacctcagca gaggtaactg gaggctgaaa ccgggtgtgt ggcctgactg gtgtgatctc 1980
 aggtcattcc agaagtggc caggaaagtc gtgagaccag gtacatgggg ggtcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtggtg gaggccaggt acatgggggg 2100
 ctacaggcact ggggtgagat aggtacacgg ggggctcagg cagggggtca gaccaggtag 2160
 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgtgtt ttcattgtag 2220
 30 ccaggctctg gcacacctgc cccaaagtcc caggaaagtg agaggccaaa gatggaggct 2280
 gcaggggctg gcggtgtggc tcacacctgt agtccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 tatgaaaaat aaaaacaaa attagctgaa catggtgtgt tgcgcctgta gttccaatac 2460
 35 ttgggaggct gaagtgggag gatcacttga gccaggagg tggaaagtgc agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtgaga gcccatctca acaacaacaa 2580
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggactt aaactacaca 2640
 ccaaagccaa gtcggtgtct cgggtgcagt gagatgagat gatgggtcct cacaccatca 2700
 cccagagacc aggggttatg caccacagg gcgggtggct cagaagggat gcgcaggacg 2760
 ttgatatacg atgacatcaa ggtgtgtcga cgaagggcag gattcatgat aagtacctgc 2820
 40 tggtagacaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880
 aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatggtgca cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actcgcacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060
 gcccattagg aaacccatgc atgtgcattc atgcacgcac acaggcaccc gtggggcccat 3120
 45 gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgtg tccgccatcc 3180
 tctcag 3186

<210> 18
 <211> 781
 50 <212> DNA
 <213> Homo sapiens

<400> 18
 55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
 ggagactgag tgaatctggg ctttaggaagt tcttaccctt tttcgcatca ggaagtgggt 120
 taaccaaac actgtcaggc tcgtctgccc gccctctcgt ggggtgagca ggcacactga 180
 tggaaaggac aggagctgtc tgggagctgc catccttccc acctgtctct gcctggggaa 240
 gcgctggggg gcctgggtct tcctgtttgc cccatgggtg gatttggggg gcctggcctc 300
 60 tctgttttgc cctgtgggtg gattgggtg tctcccgctc atggcactta gggcccttgt 360
 gcaaaccagg gccaaagggt tagggaggag ccaggcccag gctacccac ccctctcagg 420
 agcagaggcc gcgtatcacc acgacagagc cccgcgcgt cctctgtctc ccagtcaccg 480
 tcctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aacctgcggt cctgagctta acagcttcta cttctgttcc tttctgtgtt 600

17 / 18

gtggaaattt cacctggaga agccgaagaa aacatttctg tcgtgactcc tgcggtgctt 660
 gggctcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720
 gtgtctcctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttcccca 780
 g 781

5
 <210> 19
 <211> 536
 <212> DNA
 <213> Homo sapiens

10
 <400> 19
 gcaagtgtgg tgggaggcca gtgcggggccc cacctgcccc ggggtcatcc ttgaacgccc 60
 tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cgcccccggg cctgaccctg 120
 ggggctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
 15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgtgggtccc 240
 ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300
 ctgcccctgag ctectggggc cctgagcaag ttctctcccc gccccgcccgc tccagcgtca 360
 ctgggctgcc tgtctgctcg ccccgggtga ggggtgtctg tcccttctact gaggttccca 420
 ccagccaggg ccacgaggtg caggccctgc ctgcccgggc acccacacgt cctaggaggg 480
 20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gattttggcc ccgcag 536

<210> 20
 <211> 3179
 <212> DNA
 25 <213> Homo sapiens

<400> 20
 atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
 ctgtgagtga acgggggtgt ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
 30 gtctatgagt gaatgggggt gtggtcagtg cgggcccacg gcctggctgg gcctgggagg 180
 tttctgatgc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagcccca 240
 ccctggaaga cataacagta agtccaggcc cgaagggcag cagggatgct gggggcccag 300
 cttggggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggccccag 360
 ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
 35 ggctgggccc cctctctccc tgccctccac ctgcagccgt ggatccggat gtgcttccct 480
 ggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacacccat 540
 cctgtataaa atccaggatt cctctctctg aacgcccaca ctcaggttga aagtcacatt 600
 ccgctctctg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
 gtggggcagt ggaggggtgt gacacaggag gcttcagggt ggggctgggt atgctctctc 720
 40 atctctctat catctcccag tctcatctct catctctta tcactctcca gtctcatctg 780
 tcttctctct atctcccagt ctcatctgtc atctctttac catctcccag tctcatctct 840
 tatctcttta tctcttagtc tcaccagac ttacctccca gggcgggtgc caggctcaga 900
 gtggagctgg acatacgtcc ttctcaggc agaaggaaact ggaaggattg cagagaacag 960
 gagggggcggc tcagagggac gcagctcttg ggtgaagaaa cagcccctcc tcagaagtgt 1020
 45 gcttgggcca cagaaaccg agggccctgc gtgagtggct ccagagcctt ccagcaggtc 1080
 cctgttgggg ccttatggta tggccgggtc ctactgagtg caccctggac agggcttctg 1140
 gtttgagtgc agcccggacg tgcctggtgt cggggtgggg gcttatggcc actggatatg 1200
 gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260
 gggcccaagt ccacagactg tgcgtaaat gcactctggt gcctggagcc cccgtatagg 1320
 50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaaact 1380
 ggaaggggagc ggccccgggc gccgtgggag gacgacctca agtgagaggt tggacagaac 1440
 agggcggggg cttcccaggga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
 gaccaacagg tcagccattt gttcagctat ccattctcta caaagctcca gattcctgtt 1560
 tctcgggtgt ttttttgttg aaattttact caggattact tatatttttt gctaaagtat 1620
 55 tagaccctta aaaaaggat ttgctttgat atggcttaac tctaagca cctactttat 1680
 ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740
 ggtgtttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccccc aggtcaagt 1800
 gatectccgg cctcagcttc ccagagtgtc gggattacag gtgtgagcca ctgcccttgc 1860
 ctggcacttt taaaaaccac tatgtaaggt cagggtccagt ggcttccaca cctgtcatcc 1920
 60 cagtagtttg ggaagccgag gcagaaggat tgtctgaggg caggagtgtg agaccagcat 1980
 gggtaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtgggg 2040
 ccagcatctg tagtcccagc tgctcgggag gctgagtggt aggatcgctt gagcccgga 2100
 ggtcatggct gcagtgagct gtgattgtac catcgactc cagcctgggc aacagagtga 2160
 gaccctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
 65 gaaggaaaaga gaagaagaag gaagaaggaa gaagaaggaa gaaggaggcc tgctagggtc 2280

18 / 18

5 taggtagact gtcaaatctc agagcaaaat gaaaataaca aagtttttaa gggaaagaaa 2340
aaccaccagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520
gccccacgct cctgccggtc ctgcacctgc tgtaaccgtc gatgttggtg ccagggtgccc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttggtg caggcacaat tacagcccct ccccaaagat gccacgtcc ttctcctgga 2700
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcagggt gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820
10 ggggccctag aagtgagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggcctgccc 3000
acgctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
15 cagtacaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

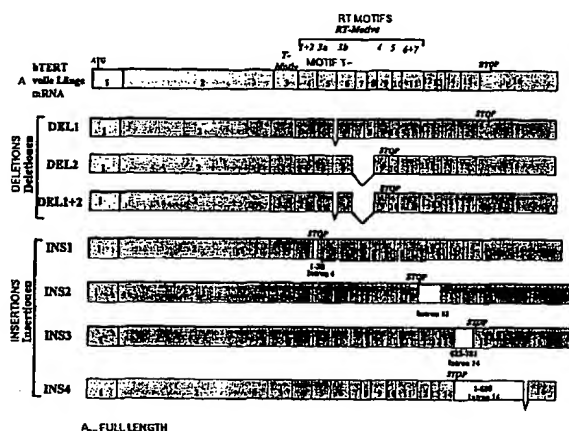


INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A3	(11) Internationale Veröffentlichungsnummer: WO 99/33998
		(43) Internationales Veröffentlichungsdatum:	8. Juli 1999 (08.07.99)
(21) Internationales Aktenzeichen: PCT/EP98/08216 (22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98) (30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE (71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE). (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE). (74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).		(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Veröffentlicht Mit internationalem Recherchenbericht. Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist. Veröffentlichung wird wiederholt falls Änderungen eintreffen. (88) Veröffentlichungsdatum des internationalen Recherchenberichts: 19. August 1999 (19.08.99)	

(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF

(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 98/08216

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/54 C12N9/12 C12N15/11 C12N15/85 C12Q1/68 A01K67/027		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12Q A01K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NAKAMURA T M ET AL: "TELOMERASE CATALYTIC SUBUNIT HOMOLOGS FROM FISSION YEAST AND HUMAN" SCIENCE, vol. 277, 15 August 1997 (1997-08-15), pages 955-959, XP002056803 ISSN: 0036-8075 figure 1A <div style="text-align: center;">--- -/--</div>	1,2
<div style="display: flex; justify-content: space-between;"> <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. </div>		
* Special categories of cited documents :		
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search <div style="text-align: center;">8 June 1999</div>		Date of mailing of the international search report <div style="text-align: center;">21/06/1999</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer <div style="text-align: center;">Andres, S</div>

1

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 98/08216

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>KILIAN ET AL: "Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types"</p> <p>HUMAN MOLECULAR GENETICS, vol. 6, no. 12, November 1997 (1997-11), pages 2011-2019, XP002086926 ISSN: 0964-6906 cited in the application page 2014, right-hand column, paragraph 3 - page 2016</p> <p>---</p>	1,2
P,X	<p>WO 98 14593 A (ANDREWS WILLIAM H ;CECH THOMAS R (US); MORIN GREGG B (US); NAKAMUR) 9 April 1998 (1998-04-09) cited in the application page 28, line 14 -, sentence 27 page 34, line 9 - page 35, line 26 page 94, line 13 - page 95, line 8 page 154, line 25 - page 155, line 5 page 217 - page 220; examples 3,4 page 266 - page 270; examples 14,15</p> <p>---</p>	1-9,11
T	<p>WICK, M. ET AL.: "Genomic organization and promoter characterization of the gene encoding the human telomerase reverse transcriptase (hTERT)"</p> <p>GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES., vol. 232, 17 May 1999 (1999-05-17), pages 97-106, XP002105207 ISSN: 0378-1119</p> <p>---</p>	
T	<p>CONG, Y.-S. ET AL.: "The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter"</p> <p>HUMAN MOLECULAR GENETICS., vol. 8, 1 January 1999 (1999-01-01), pages 137-142, XP002105208 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0964-6906</p> <p>---</p>	
T	<p>TAKAKURA, M. ET AL.: "Cloning of human telomerase catalytic subunit (hTERT) gene promoter and identification of proximal core promoter sequences essential for transcriptional activation in immortalized and cancer cells"</p> <p>CANCER RESEARCH., vol. 59, 1 February 1999 (1999-02-01), pages 551-557, XP002105209 ISSN: 0008-5472</p> <p>---</p> <p style="text-align: center;">-/--</p>	

INTERNATIONAL SEARCH REPORT

Inter. onal Application No

PCT/EP 98/08216

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	<p>HORIKAWA, I. ET AL.: "Cloning and characterization of the promoter region of human telomerase reverse transcriptase gene"</p> <p>CANCER RESEARCH.,</p> <p>vol. 59, 15 February 1999 (1999-02-15),</p> <p>pages 826-830, XP002105210</p> <p>AMERICAN ASSOCIATION FOR CANCER RESEARCH,</p> <p>BALTIMORE, MD., US</p> <p>ISSN: 0008-5472</p> <p>-----</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 98/08216

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9814593 A	09-04-1998	AU 4803697 A	24-04-1998
		AU 4807397 A	24-04-1998
		DE 19743497 A	20-08-1998
		DE 841396 T	24-09-1998
		EP 0841396 A	13-05-1998
		FI 990655 A	24-03-1999
		FR 2757177 A	19-06-1998
		GB 2317891 A,B	08-04-1998
		GB 2321642 A	05-08-1998
		JP 10234384 A	08-09-1998
		WO 9814592 A	09-04-1998

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 98/08216

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES IPK 6 C12N15/54 C12N9/12 C12N15/11 C12N15/85 C12Q1/68 A01K67/027		
Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK		
B. RECHERCHIERTE GEBIETE Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole) IPK 6 C12N C12Q A01K		
Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	NAKAMURA T M ET AL: "TELOMERASE CATALYTIC SUBUNIT HOMOLOGS FROM FISSION YEAST AND HUMAN" SCIENCE, Bd. 277, 15. August 1997 (1997-08-15), Seiten 955-959, XP002056803 ISSN: 0036-8075 Abbildung 1A --- -/--	1,2
<div style="display: flex; justify-content: space-between;"> <div> <input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen </div> <div> <input checked="" type="checkbox"/> Siehe Anhang Patentfamilie </div> </div>		
<div style="display: flex;"> <div style="flex: 1;"> <p>* Besondere Kategorien von angegebenen Veröffentlichungen :</p> <p>"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist</p> <p>"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist</p> <p>"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)</p> <p>"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht</p> <p>"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist</p> </div> <div style="flex: 1;"> <p>"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist</p> <p>"X" Veröffentlichung von besonderer Bedeutung: die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden</p> <p>"Y" Veröffentlichung von besonderer Bedeutung: die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist</p> <p>"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist</p> </div> </div>		
Datum des Abschlusses der internationalen Recherche		Absendedatum des internationalen Recherchenberichts
8. Juni 1999		21/06/1999
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Bevollmächtigter Beauftragter Andres, S

1

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 98/08216

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	KILIAN ET AL.: "Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types" HUMAN MOLECULAR GENETICS, Bd. 6, Nr. 12, November 1997 (1997-11), Seiten 2011-2019, XP002086926 ISSN: 0964-6906 in der Anmeldung erwähnt Seite 2014, rechte Spalte, Absatz 3 - Seite 2016 ---	1,2
P,X	WO 98 14593 A (ANDREWS WILLIAM H ;CECH THOMAS R (US); MORIN GREGG B (US); NAKAMUR) 9. April 1998 (1998-04-09) in der Anmeldung erwähnt Seite 28, Zeile 14 -, Satz 27 Seite 34, Zeile 9 - Seite 35, Zeile 26 Seite 94, Zeile 13 - Seite 95, Zeile 8 Seite 154, Zeile 25 - Seite 155, Zeile 5 Seite 217 - Seite 220; Beispiele 3,4 Seite 266 - Seite 270; Beispiele 14,15 ---	1-9,11
T	WICK, M. ET AL.: "Genomic organization and promoter characterization of the gene encoding the human telomerase reverse transcriptase (hTERT)" GENE: AN INTERNATIONAL JOURNAL ON GENES AND GENOMES., Bd. 232, 17. Mai 1999 (1999-05-17), Seiten 97-106, XP002105207 ISSN: 0378-1119 ---	
T	CONG, Y.-S. ET AL.: "The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter" HUMAN MOLECULAR GENETICS., Bd. 8, 1. Januar 1999 (1999-01-01), Seiten 137-142, XP002105208 OXFORD UNIVERSITY PRESS, SURREY., GB ISSN: 0964-6906 ---	
T	TAKAKURA, M. ET AL.: "Cloning of human telomerase catalytic subunit (hTERT) gene promoter and identification of proximal core promoter sequences essential for transcriptional activation in immortalized and cancer cells" CANCER RESEARCH., Bd. 59, 1. Februar 1999 (1999-02-01), Seiten 551-557, XP002105209 ISSN: 0008-5472 --- -/--	

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 98/08216

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
T	<p>HORIKAWA, I. ET AL.: "Cloning and characterization of the promoter region of human telomerase reverse transcriptase gene"</p> <p>CANCER RESEARCH.,</p> <p>Bd. 59, 15. Februar 1999 (1999-02-15),</p> <p>Seiten 826-830, XP002105210</p> <p>AMERICAN ASSOCIATION FOR CANCER RESEARCH,</p> <p>BALTIMORE, MD., US</p> <p>ISSN: 0008-5472</p> <p>-----</p>	

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP 98/08216

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
WO 9814593 A	09-04-1998	AU 4803697 A	24-04-1998
		AU 4807397 A	24-04-1998
		DE 19743497 A	20-08-1998
		DE 841396 T	24-09-1998
		EP 0841396 A	13-05-1998
		FI 990655 A	24-03-1999
		FR 2757177 A	19-06-1998
		GB 2317891 A, B	08-04-1998
		GB 2321642 A	05-08-1998
		JP 10234384 A	08-09-1998
		WO 9814592 A	09-04-1998
